CONTENTS OF VOLUME 1

ASHBY, W. Ross: The Effect of Experience on a Determinate Dynamic System, 35, (Jan)

BATESON, G., JACKSON, D., HALEY, J., AND WEAKLAND J.: Toward a Theory of Schizophrenia, 251, (Oct)

CAMPBELL, DONALD T.: Adaptive Behavior from Random Response, 105, (Apr)

EASTON, DAVID: Limits of the Equilibrium Model in Social Research, 96, (Apr)

GERARD, R. W., KLUCKHOHN, C., AND RAPOPORT, A.: Biological and Cultural Evolution: Some Analogies and Explorations, 6, (Jan)

HALEY, J.: See BATESON, G., jt. author

JACKSON, D.: See BATESON, G., jt. author

KLUCKHOHN, CLYDE: See GERARD, R. W., jt. author

KOCHEN, MANFRED: The Logical Nature of an Action Scheme, 265, (Oct)

LEVY, MARION J., JR.: See KOCHEN, MANFRED, jt. author

LUCE, R. DUNCAN AND ROGOW, ARNOLD A.: A Game Theoretic Analysis of Congressional Power Distributions for a Two-Party System, 83, (Apr)

MANN, RICHARD D.: A Critique of P. E. Meehl's *Clinical Versus Statistical Prediction*, 224, (July)

MEIER, RICHARD L.: Communications and Social Change, 43, (Jan)

MILLER, JAMES G.: Review of *Toward a Unified Theory of Human Behavior*, 319, (Oct)

MONEY, JOHN: Mind-Body Dualism and the Unity of Bodymind, 212, (July)

OSGOOD, CHARLES D.: Behavior Theory and the Social Sciences, 167, (July)

PERRY, STEWART E.: Observations on Social Processes in Psychiatric Research, 290, (Oct)

RAPOPORT, A.: See GERARD, R. W., jt. author

RAPOPORT, A.: A Critique of Frederick Mosteller and Robert R. Bush's *Stochastic Models for Learning*, 59, (Jan)

RAPOPORT, A.: Review of *Problems in Human Communication and Control*, 309, (Oct)

ROGOW, ARNOLD A.: See LUCE, R. DUNCAN, jt author

RUFF, GEORGE E.: A Critique of Kenneth Mark Colby's *Energy and Structure in Psychoanalysis*, 143, (Apr)

SMITH, NICHOLAS, M., JR.: A Calculus for Ethics: A Theory of the Structure of Value. Part I, 111, (Apr)

SMITH, NICHOLAS M., JR. A Calculus for Ethics: A Theory of the Structure of Value. Part II, 186, (July)

TRUCCO, ERNESTO: Review of *Information Theory in Psychology; Problems and Methods*, 315, (Oct)

TUMIN, MELVIN M.: Some Disfunctions of Institutional Imbalance, 218, (July)

WEAKLAND, J.: See BATESON, G., jt author

EDITORIAL

BEHAVIORAL SCIENCE, A NEW JOURNAL

THE remarkable growth of interdisciplinary interest in behavioral science during the last decade is the fundamental justification for this new periodical. Even though its Board of Editors knows that there are many scholarly journals, we believe that there is a clear need for a publication of the sort we have planned. It is no news to behavioral scientists that the last ten years have witnessed at least as much ferment in the disciplines concerned with behavior, normal and abnormal, individual and group, as in any other field. And this is as it must be if man's understanding of himself is to advance along with his accelerating comprehension of the physical universe. The rise of natural science and the flood of its applications have been paced by the creation of broad theories. It is to the development of such theories of behavior and to their empirical testing that this publication is dedicated.

Perhaps it is meaningful to identify swings in the history of ideas. If so, we may at present discern a pendular motion in the behavioral sciences from massive concentration on empiricism for its own sake toward systematization of empirical insights. It is likely that this area of science, like the others, will thrive best on an effective union of observation and theory.

For readers who are not themselves expert in the sciences of man, it may be desirable to sketch their place in the world of today. As history shows, man has for centuries turned his energies to exploring and exploiting nature. Always refining his methods of learning, always seeking ways to cope with the forces of nature, he has developed the tools and machines that have given us the whole complex, civilized world of today. This he has accomplished by the scientific method—the process of carefully observing what happens, trying to explain how it happens, and testing the explanation by predicting what will happen next. Then have come attempts to control the forces that have been recognized. Almost carelessly reversing the flow of rivers, bringing relief from mass epidemics, accomplishing mass destruction, even attempting to change the weather and planning to leave his planet, mankind has been incredibly successful at solving the riddles of the material world.

Man's most baffling enigma remains, as it has always been, himself. He has been unable to fathom with any precision those laws of human nature which can produce social inequality, industrial strife, marital disharmony, juvenile delinquency, mental illness, war, and other widespread miseries.

Many different approaches have been used in the study of behavior—

mathematical biology, biochemistry, physiology, genetics, medicine, psychiatry, psychology, sociology, economics, politics, anthropology, history, philosophy, and others. Though the term "interdisciplinary" is widely current, and for a long time efforts at collaboration have been made, true unification of these fields still remains an unattained goal. And within each are various schools. Their approaches and skills are specific, but the problems are general. Can the scientific method solve the larger, more pervasive questions about man as well as the smaller, more particular ones? Is the tool with which man has won his victories over the physical world applicable to uncovering the laws which govern man's conduct, the deepest causes of our strife and our harmony? If the fragments of multiple sciences were brought together in a unitary behavioral science and all the separate skills focused on the study of human behavior, perhaps the time required to find answers to these questions could be reduced. It is possible that inadequacies in the present studies of man could thus be avoided. The uniformities among disciplines could be recognized; better communication among them established; generality of findings magnified; additional benefits derived from comparing theories in diverse fields, explaining both similarities and differences; and the validity and applicability of empirical work increased by planning individual studies as components of an explicit mosaic of research strategy.

About 1949 a group of faculty members at the University of Chicago, some of whom have now moved to the University of Michigan, began to consider whether a sufficient body of facts exists to justify developing empirically testable general theories of behavior. This group used the term "behavioral science" to cover the diverse areas of their interests, primarily because its neutral character made it acceptable to both social and biological scientists.

Most of the participants were at first skeptical that our comprehension of these different areas had advanced sufficiently to justify such activity. The first meetings engendered a general hopelessness as the diversity of languages and the multitude of approaches to the study of man became increasingly apparent. But then we began to see among us certain commonalties of thinking, despite their many linguistic disguises, and this agreement gave us hope that our efforts were not unrealistic.

Members of this group have met intensively for several years as the Committee on Behavioral Science at the University of Chicago. Some are continuing this activity at Chicago; others went to staff the new Mental Health Research Institute, established in August, 1955, at the University of Michigan; and there they were joined by still others. The Regents of the University and the Legislature of the State of Michigan established this Institute on a permanent basis.

The aim is to conduct basic research; the expectation, that from such

research will flow contributions, particularly in the field of mental health and disease, that will help to solve the many problems of human relations. Our understanding of mental illness is primitive compared with our knowledge of other forms of disease, partly because of the complexity of the problems and partly because research efforts have not been commensurate with their magnitude. Public interest in these issues is growing rapidly, as evidenced by the new or greatly increased appropriations for investigation by state legislatures and the Congress, and by additional support from foundations.

In this area of behavioral science there are numerous schools with conflicting beliefs. No one as yet has seen how the insights of psychodynamics, the projective techniques of psychology, the facts of neuropathology, the discoveries of endocrinology, biochemistry, and neurophysiology, and the concepts of social science can be merged into a single framework for explaining the biological and psychiatric and social phenomena of mental illness. There is need now for renewed and exhaustive examination of these separate matters, and for creative attempts to integrate them, to test them empirically, and to apply them.

Such studies should be carried out at various levels. Our present thinking—which may alter with time—is that a general theory will deal with structural and behavioral properties of *systems*. The diversity of systems is great. The molecule, the cell, the organ, the individual, the group, the society are all examples of systems. Besides differing in the level of organization, systems differ in many other crucial respects. They may be living, nonliving, or mixed; material or conceptual; and so forth.

The strategy of the Michigan Institute's work will emphasize identification of general principles, which extend across various levels of systems. We shall attempt to clarify and make precise both the general principles and the particular differences; and to test—in laboratories and in clinics, by group studies and by social surveys, with whatever methods prove appropriate—the validity and usefulness of such analysis. Research techniques will probably be derived from several areas, including the physiological, psychological, economic, political, social and cultural.

Although the Institute expects to pay particular attention to the similarities and dissimilarities among different behaving systems, this is only one of many legitimate approaches to behavior theory. *Behavioral Science*, as a journal with wider scope than any single Institute, will welcome articles which are constructively critical of this orientation or which advance other alternative strategies, as well as articles which present relevant empirical studies.

This is the official publication of the Mental Health Research Institute at the University of Michigan. As such it will contain edited records of

roundtable discussions on theory and reports of other activities involving the Institute. It is hoped that Ann Arbor can in the summer offer its facilities as a meeting center for scientists, many from other institutions, who are concerned with behavior theory or mental health or with related experimental and clinical work. Reports of such conferences and workshops will also be included in this journal.

Other centers are carrying out closely related work. The Committee on Behavioral Science at Chicago, for example, maintains its original interests, and other universities are supporting or planning comparable programs. A particularly significant focus of activity is the Center for Advanced Study in the Behavioral Sciences established by the Ford Foundation and located at Stanford, California. This journal will welcome contributions from scholars at these centers or elsewhere. It should serve as one channel of communication for members of the ever-increasing group engaged in advancing the sciences of man.

We are aware of no present journal with a primary policy of making its pages available to representatives of any field—the humanities, the social sciences, the biological and medical sciences, and the physical sciences—to discuss theory concerning behavior, and empirical studies clearly oriented to such theory. It has been rare for physicists, psychiatrists, political scientists, and historians to publish in, or even read, the same journal. We shall strive to achieve this end.

What kind of article will be acceptable to *Behavioral Science*? Length will not be controlling; long articles will be published if the content justifies them, but a range from 3,000 to 20,000 words is suggested, with 10,000 as the mode. Preference will be given to articles written in a style understandable to the "layman-scientist" interested in behavior. Indeed, the diversity of science has now become so vast that the historian is in effect a layman to writings of a physiologist, and the mathematician to those of a political scientist. The language of common scholarship must, therefore, today be the Esperanto of behavioral science. All major articles will have a brief introductory summary indicating the general tenor of the article and its significance for behavioral science.

Some technical material will also be printed, particularly in the sections concerned with critical analysis of important books or with abstracts of recent publications. The summaries of such articles, however, will be written in a nontechnical fashion. It is the intention of the editors that mathematical statements and arguments will be, insofar as possible, restated and explained in such a way that they will be meaningful to those not fluent in the mathematical language.

Readers are encouraged to submit manuscripts; to comment on and criti-

cize, perhaps for publication, material appearing in the journal; and, if they be so moved, to make suggestions concerning policies.

FRANZ ALEXANDER
ALEX BAVELAS
DAVID EASTON
RALPH W. GERARD
DONALD G. MARQUIS
JACOB MARSCHAK
JAMES G. MILLER
ANATOL RAPOPORT
RALPH W. TYLER
RAYMOND W. WAGGONER

Biological and Cultural Evolution

Some Analogies and Explorations

by R. W. Gerard, Clyde Kluckhohn, and Anatol Rapoport

From the Center for Advanced Study in the Behavioral Sciences; the Mental Health Research Institute, University of Michigan; and the Department of Anthropology, Harvard University

Since Darwin there has been much discussion pro and con as to whether profitable analogies can be drawn between the evolution of species and the development of different kinds of primitive and advanced human societies. This question is reviewed in this article by an interdisciplinary team. It is suggested that orientations and methods which have been employed to investigate biological evolution might also be used in the study of the evolution of society or culture. Perhaps these will throw light on the theoretical problem of the similarities and differences in the two sorts of evolution. Experiments are suggested in which small groups work together on problems under special constraints of customs and languages dictated by the experimenters. In addition to this experimental approach to the study of different cultures, and normal and abnormal behavior in them, the historical method of comparing biological and social evolution is employed in an examination of language.

INTRODUCTION: ON GENERALIZING

THIS paper gives a preliminary report on work initiated by an interdisciplinary group (anthropology, biology, mathematics, social psychology, and sociology were represented[1]) at the Center for Advanced Study in the Behavioral Sciences during 1954–1955. We are quite conscious of the tentativeness and indeed the controversial nature of many of our suggestions; but are nevertheless persuaded that it may be useful both to our colleagues and to further work planned by us to share these points of view and some provisional findings.

The group began with little more than a general conviction that the question of similarities between biological and cultural change deserved re-examination from a somewhat fresh angle.[2] After some early unstructured discussion we decided that the carefully worked-out knowledge of population genetics might provide the most suggestive model from biology. We then turned to linguistics as representing the area in the cultural realm where—at least in phonology—fundamental units organized into a system have been firmly established, and examined phonological and phonemic theory and empirical data on phonetic evolution, searching for similarities with the processes demonstrated in Mendelian genetics. The third topic in our program, that of constructing experimental analogues to some aspects of the cultural process, is undoubtedly the most innovating development of the work group. How it arose out of group discussion and how its construction was influenced by consideration of the first two topics would become fully evident only to someone who read a complete record of our sessions. We believe, however, that careful readers of the

[1] Besides the present authors, Alex Bavelas, R. Duncan Luce, and Louis Schneider participated in the group discussions. Bavelas contributed the key ideas of the experimental outline described below and took the lead in working them out. Only his modesty prevents his co-authorship.

[2] See, e.g., Gerard (5, 6).

present communication will catch interconnections and implications between the three topics which we cannot take the space to spell out in full, explicit detail.

Before reporting seriatim on our three major topics, it seems important to state explicitly two points of view which emerged in present form during our sessions and which guided our investigations. The first relates to the role of analogy in science. The second concerns our particular conception of interdisciplinary research.

The role of analogy in science

There is a certain organized necessity about generalizing or thinking across fields because all knowledge, indeed perhaps all perception, starts with the recognition of similarities. One advances, to the degree that the data make possible, from similarities to analogies to congruences to identities. Starting with a similarity between sound waves and water waves, or even between heat flow and electric flow, there is progressive likeness, to the congruence between light waves and radio waves (which differ only in wave length, as do red and orange). One may instance comparable chemical examples: rock and soil are similar; alcohols and other organic compounds are analogous; members of a series of alcohols are congruent; finally, the temporally interchangeable forms of a molecule, as in the keto-enol shift, are two tautologies or different forms of an identity and are actually called tautomers. The mathematical and logical tautologies which simply state a formal isomorphism between the variables of different phenomena—as the identical equations for mechanical and electric oscillators—carry formal identities across widely disparate content areas.

If by looking across the board, one achieves a new approach, a new way of viewing something and new insights result, there is a tremendous gain. If the particular universe under examination suddenly becomes part of a wider universe, if generalization can be extended to include new kinds of elements, new power results. In the behavioral sciences, in particular, there is need to be able, as it were, to add peanuts and elephants. Since arithmetic is not applicable, it is useful to generalize from simple numerical to abstract logical relations. In the latter is included set theory, for example, which does permit one to handle peanuts and elephants together.

Curiously, two extreme attitudes toward analogy both stem from a recognition of the nature of deductive method. One is epitomized in the well-known dictum, "Proof by analogy is not proof." The other is contained in a statement of the mathematician, E. H. Moore, who maintained that any similarity between two mathematical theories implies the existence of a more general theory of which the two are special cases. Stated in terms of formal logic, the first statement is a recognition of the fallacy of the all too common *non sequitur*: A implies B; A and C have elements in common; therefore C implies B. The second statement is, on the face of it, trivial. All it says is that if sets A and B have elements in common, there is a set C which contains them both. Set C can always be constructed so as to contain both A and B. However, self-evident as are both strictures from formal logic (or the equivalent algebra of sets), they are philosophically significant when applied to theory construction because they point both to the fruitfulness and the shortcomings of analogy.

To take an example from the history of physiology, the first tissue intensively studied for its chemical components and metabolic events was skeletal muscle, being the easiest to handle. Here it was found that during activity, sugar splits into lactic acid, which is partly burned to carbon dioxide, that certain phosphate compounds break down, etc. When work later extended to other tissue, as brain, liver, kidney, the same things were looked for and, to be sure, mostly found. Here the analogy of one tissue to another was useful and led to the expectation of common features in all living tissue. However, many workers rather thoughtlessly came to expect that all the chemistry of muscle would apply to, say, brain. This expectation is obviously unjustified, for if brain had exactly the same metabolism as muscle the two tissues would be identical. There must be some point at which similar metabolisms begin to diverge; and the divergences

will ultimately teach us even more about tissues than the similarities. Nevertheless, great progress can be made by starting with the similarities.

Let us now look at the second statement, which, we said, was logically trivial, because in the realm of abstract sets (with which all purely deductive thinking is exclusively concerned) a set can always be constructed to include any two given sets. For instance, one could take the set "objects" to include the sets "peanuts" and "elephants." The task, however, is not merely to construct any theory (set) from which two given theories can be deduced as special cases (which includes two given sets), but rather to construct a simple and coherent theory (an easily recognizable set) which has this property. Thus "animals and their food" includes "peanuts and elephants" more significantly than "object"; "oxidation" relates to "respiration" and "combustion" more fruitfully than "process."

Indeed, the history of theoretical physics is the most impressive example of just such constructions. A linear mechanical oscillator behaves like an alternating current circuit, where the impressed potential difference of the latter corresponds to the impressed force of the former, the current to velocity, charge to displacement, ohmic resistance to frictional resistance, capacitance to the rigidity modulus, and inductance to mass. Different as the respective contents of the situations are, a single differential equation governs the behavior of both. The recognition of the unity of heat and mechanical energy, of light and electromagnetic waves, of mass and energy, all triumphs of theoretical physics, are directly traceable to the persistent search for general schemata to embody phenomena which exhibit analogous features.

In chemistry, the classification of compounds according to analogous properties led to the discovery of the primary classical building blocks of matter, the elements, so that analogous chemical properties of substances were revealed as manifestations of analogous structures of their respective molecules. Moreover, the classification of the elements themselves, according to their analogous properties, led to the construction of the periodic table and subsequently to the discovery of the structure of atoms.

In biology, one distinguishes between analogies (here "analogy" has a more restrictive meaning) and homologies of anatomical structures. Both concepts are exceedingly fruitful. The first points to similarities of adaptation to environment, the second to phylogenetic origins.

There is no question, then, of the fruitfulness of analogical thinking if it is pursued, not with the view of "proving" that two similar things are instances of the same thing, but with the view of inquiring what may be behind the similarity. At the same time, the tautological nature of Moore's "existence conjecture" should serve as a warning against letting a powerful generalization degenerate into an empty generality. It is probably of little consequence that key holes and tunnels are both "orifices" and that the Washington Monument and a pencil are both "oblong objects" (although both sets may appear in certain contexts in the significant set, phallic symbols). One should guard against the temptation of undisciplined and vague analogizing offered by such powerful general concepts as energy, homeostasis, stress, interaction, structure, system, libido, etc. Recent advances in the mathematical theory of communication and in cybernetics have introduced the terms: information, entropy, noise, channel capacity, feedback, servo-mechanism, stability, time series, etc., etc. Although defined with mathematical precision in their technical context, these terms are immensely suggestive of a vast variety of situations. Accordingly, well-meaning but often insufficiently informed enthusiasts, under the impression that the use of these terms renders metaphorical speculations "scientific," tend to use them promiscuously and thereby often obscure their real value.

In summary, we hold the recognition of analogies to be fundamentally useful in the construction of theory, the usefulness depending on the precision with which analogical entities and relations in two or more situations can be identified and on the nature of the questions which arise as a conse-

quence of the comparison. Thus, the spread of excitation in a neural net, the spread of an epidemic, and the spread of information in a population are usefully analogous, because in all three cases we have (*a*) identifiable carriers (neurons and individuals), (*b*) identifiable relations among the carriers relevant to the spread (existence or nonexistence of a connection or contact), (*c*) an identifiable state (excited, infected, informed, and their opposites), and (*d*) nonconservative transfer of the state (extinction and proliferation possible in all cases). The analogy is fruitful to the extent that it is possible to treat all phenomena by equations of a similar type and to the extent that the parameters of the equations suggest further analogies, e.g., refractory state of a neuron and acquired immunity to a disease, frequency distribution of synaptic delays and the time course of infectivity, etc.[3] The ultimate fruitfulness of the analogy depends, of course, on extensive testing of the proposed equations under a variety of conditions and on the derivation through the observed divergences among the phenomena of the features characteristic for each phenomenon. Analogical thinking is thus in our view not so much a source of answers on the nature of phenomena as a source of challenging questions.

Interdisciplinary research

Our second point of departure, the interdisciplinary, stems to considerable degree from our faith in the fruitfulness of analogical thinking, properly controlled. The example just given connects, by apparent similarity of underlying structure, events which content-wise neurologists, epidemiologists, and sociologists had appropriated for their special studies. But, again, we hold no simplistic brief for all research dubbed "interdisciplinary." There is no dearth of conferences and symposia nowadays in which many disciplines are "represented." The joint outputs of such enterprises fall short of their mark, however, if they are little more than "mosaics" of gems, each carefully polished by its appropriate expert. The results of joint efforts should be more than additive or else the effort involved in talking across disciplinary boundaries is hardly justified. In the present instance, while no one of us would pretend to control of an additional field, we worked together long enough and intensively enough so that we think at least we all have some understanding of one another's biases. While the original drafting of the sections of this paper followed disciplinary lines to some extent, there is now scarcely a page which has not been reworked from the angle of at least one additional subject.

In concluding this introduction, we present a first sketch of our thematic analogy. The anthropological study of culture has shifted emphasis from the totality of artifacts and directly observable customs to the concepts and premises distinctive of each culture. These are held to constitute a system of least common denominators within a culture, of which the observables are the expressions. Thus it is possible to isolate more fundamental and therefore more stable units of analysis. Such units show empirically a remarkable stability, in spite of the fact that an entire population of culture-carriers is replaced within a few decades. On the other hand, cultures certainly change through time. These changes are in many respects systematic and cumulative. In other ways they resemble the "spontaneous" or "accidental" nature of genetic mutations. As a preliminary scheme of our basic analogy, we offer the suggested correspondences, shown in Table 1.

A major difference between species and cultures may rest in the relative biological isolation of breeding populations as against the extensive "cross-breeding" continually observed in the history of cultures. However, at least in the case of Homo sapiens, gene exchange among human beings has occurred on a scale comparable to that demonstrated between cultures and subcultures. Also, just as new species may result from isolation and adaptations of breeding populations, so new cultures evolve when subcultures develop through time in varying environments, either in relative isolation or in contact with cultures of different types than those which influenced the original culture that becomes

[3] See, e.g., Rapoport (19, 20) and Landau and Rapoport (16).

TABLE I

CORRESPONDENCES BETWEEN BIOLOGICAL AND CULTURAL EVOLUTION

Biological Evolution	Cultural Evolution
Distinct species and varieties	Distinct cultures and subcultures
Morphology, structural organization	Directly observable artifacts and customs distinctive of cultures
Physiology, functional attributes	Functional properties attributable to directly observable cultural characteristics.
Genetic complex determining structures and functions	"Implicit culture"*—i.e. the inferred cultural structure or "cultural genotype"
Preservation of species but replacement of individuals	Preservation of cultures but replacement of individuals and artifacts
Hereditary transmission of genetic complex, generating particular species	"Hereditary" transmission of idea-custom-artifact complexes, generating particular cultures
Modification of genetic complex by mutations, selection, migration, and "genetic drift"	Culture change through invention and discovery; adaptation; diffusion and other forms of culture contact; "cultural drift"
Natural selection of genetic complexes generally leading to adaptation to environment	Adaptive and "accidental" (i.e. historically determined) selection of ideas, customs, and artifacts
Extinction of the maladapted and maladjusted species	Extinction of maladapted and maladjusted cultures

*For an explanation of this concept, see Kluckhohn (15).

TABLE 2

CORRESPONDENCES BETWEEN CULTURES OR SUBCULTURES AND SUBSPECIES OF A SINGLE SPECIES

Culture or Subculture	Subspecies
Partial isolation of subspecies—"cellulation"	Partial isolation of cultures—"cellulation"
Cross-breeding through migration and limited interbreeding	"Cross-breeding" through diffusion of ideas, customs, artifacts
Hybrid vigor	Hybrid vigor?

modified. It would be possible to extend the analogy by comparing both cultures and subcultures to subspecies of a single species. We have then the correspondences shown in Table 2.

We suspect further that the "inner logic" of the organism is paralleled by the "inner logic" of a culture. The survival potential of a given biological characteristic must be evaluated not only in terms of its use in a certain external environment but also of the internal organization of the organism. The incisors of a horse would be of little use to a tiger, whose habits are attuned to hunting animal prey; fleetness is hardly called for in a tapeworm nourished in a pig's intestinal tract. Similar considerations apply to the fitness of a particular cultural trait in each distinctive cultural pattern. On the other hand, organisms and cultures may be "ready" for radical innovations so that when these come on the scene through biological or cultural mutation or diffusion, the evolution, whether of an organism or of a culture, may be greatly accelerated.

There are both historical and experimental models for the study of evolution. Earlier biological work was largely historical, but genetics now has an experimental base. Linguistics also has an experimental side. But we shall use linguistics as our historical analogue, returning in our final section to some suggestions for possible experimental procedures for investigating at least some essential particulars of the evolution of cultures.

GENETIC EVOLUTION

A great body of data exists concerning the evolution of biological groups, not only describing the phenomena of organic evolution but also analyzing the mechanisms. Organic evolution depends on the chemical structure and action of genes, including their arrangement; species appear and change as deter-

mined by the shifts in the gene pool of the population. We shall, therefore, first examine the gene and factors influencing the gene pool, and then look for cultural units and processes that may prove comparable.

The evolutionary factors to be considered are mutation, selection, migration, and genetic drift. Together they determine the genetic characteristics of a population, namely the genetic complex and the genetic distribution.

The genetic complex. The gene, a clearly defined unit of biological inheritance, was initially a hypothetical construct, invented to fit the results of breeding experiments, but is now a practically identified nucleo-protein molecule. A given gene may appear in the germinal cell from which the organism develops (a biparental zygote or a uniparental spore), in one of several variations, called alleles, and is then reproduced in that form in every cell of the organism (except for occasional "somatic mutations," which will not be considered here). A gene may be altogether absent from the germinal cell or represented by a completely inert allele. To subsume all cases, we shall speak of the presence or absence of a particular allele. An allele, then, is present or absent as an indivisible unit in an organism and is passed on to offspring in this all-or-none manner. The presence or absence of a given allele in a given individual is ultimately independent, when linkage and crossing over operate, of the presence or absence of other alleles in that individual. The genes interact with each other and with their environment, ranging from intracellular to extraorganismic, to produce an organism. The action is through a large series of steps, each dependent on the preceding and the current situation. The genes act largely as enzymes to catalyze the steps. The organism will thus be largely, though not entirely, determined by the particular combination of alleles present in the spore or zygote.

Since the number of genes is of the order of thousands for most organisms (20,000 to 40,000 for man) and there are commonly several alleles possible for each gene, the number of possible combinations is prodigious. With only two alleles and a thousand genes, this number is 2^{1000}, incomparably greater than the surmised number of atoms in the universe. A particular combination of alleles which characterizes an individual is called the *genotype*. The set of external characteristics determined by the genes and their interaction with the environment is called the *phenotype*. The totality of genotypes of a population is called its *genetic complex*. Just as the "fate" of the individual is to a large extent determined by its genotype, so the evolutionary "fate" of a population is to large extent determined by what happens to its genetic complex. Note, however, that the genotype of an individual is normally set in the zygote or spore, and subsequent mutations in the individual are unlikely to be of importance with respect to his phenotype (some somatic mutations, especially pathological ones, excepted); while the genetic complex of a population, on the other hand, may change drastically even in the absence of mutations, since it depends on the proportions within it of the various alleles, which proportions can change radically through differential selection of genotypes from generation to generation.

The genetic distribution. Consider each gene as either present or absent in an individual. This can be done, even when there are many alleles (there are, for example, over 80 at the B locus "red cell antigen" in cattle), by simply considering the gene as represented by one of its alleles, the so-called normal or "wild type," so that the presence of any other allele or the absence of the gene altogether is counted as an absence. Having thus defined a gene as either present or absent, in an individual, one can define the frequency of the gene in a population by the fraction of the population in which the gene is present. In general, this fraction q will vary from generation to generation. The several factors contributing to the variations fall into two classes: The *systematic drift* includes changes in frequency due to mutation, selection, migration, and other "pressures," which can be considered constant over many generations; the *random drift* includes the random fluctuations in the coefficients which characterize the systematic

drift and, especially, *sampling variance* (see below).

Over many generations, a gene will be represented in a given fraction of the population with a certain relative frequency. One can thus speak of the "probability frequency distribution of the gene frequencies." This distribution of gene frequencies is a mathematical function, whose independent variable q is a fraction of the population (ranging, of course, from 0 to 1), and whose dependent variable $\varphi(q)$ is the probability that in a given generation the gene in question is represented in the fraction q of the population.

It is shown in the mathematical theory of evolution[4] that the distribution of gene frequencies is given by the following expression:

$$\varphi(q) = \frac{C}{\sigma^2_{\delta_q}} \exp\left\{2 \int^q \frac{\Delta q}{\sigma^2_{\delta_q}}\, dq\right\}. \qquad [1]$$

Here Δq is a function of q representing the effect of the systematic drift factors (selection, migration, and mutation pressures) and $\sigma^2_{\delta_q}$ is the variance in q representing the random drift factors. The lower limit of q in the integral is immaterial, since it is absorbed in the normalizing constant C, chosen so as to make $\int_0^1 \varphi(q)\, dq = 1$.

The form of this frequency distribution indicates how the representation of the gene in question varies from generation to generation, and so the character of the function $\varphi(q)$ indicates the variability in time of the genetic complex of the population. Thus, when the distributions of all genes are strongly bunched around their respective average values, the representation of each gene remains almost constant from generation to generation, that is, the genetic complex remains constant. When, on the other hand, the individual distribution functions are spread out (in the extreme case being bunched around the values $q = 0$ and $q = 1$) the representation of individual genes fluctuates wildly, and so the genetic complex undergoes large fluctuations. As we shall see, the evolutionary fate of a population depends on the extent and the character of these fluctuations. The latter depend in specific ways upon the parameters of the gene frequency distributions Δq and $\sigma^2_{\delta_q}$, which, in turn, depend on the above-mentioned evolutionary factors and on population size. This chain of dependence constitutes the current theory of population genetics. Let us consider the evolutionary factors in turn.

Mutation. Genes mutate, that is, pass abruptly from one allele to another. On the basis of existing evidence, the overwhelmingly dominant view is that mutations occur at random. Certain environmental factors may change the rates of mutation, but no systematic correlations have been unequivocally established between environmental factors and the direction of the changes. Each gene has certain constant probabilities of mutating from one of its alleles to each of the others, which probabilities may, however, depend on the presence of other alleles. If a gene has a probability u of mutating in a generation, and if the probability of mutating back to the original allele is v, the contribution of mutation pressure to the parameter Δq will be

$$-uq + v(1 - q) \qquad [2]$$

Selection. Each genotype, as manifested in its phenotype, is characterized by a "survival potential," a reflection of the degree of adaptation to environment and of fecundity. Greater survival potential or survival advantage of a genotype is manifested in its increased representation in the next generation, because of its favorable differential in birth and death rates as compared to other genotypes. No matter how slight (numerically) is the relative increase of representation, eventually the genotypes possessing such advantage would inevitably replace the others, were it not for the operation of such other factors as mutation and migration pressures, which insure the replenishment of deviant genes in the population. Since selection works on individuals (actually on larger units also) rather than on genes, any selection coefficient should be a function of the entire genotype. For simplicity, however, a selection coefficient has been taken to

[4] The mathematical theory presented here has been developed largely by Sewall Wright (14, 26, 27).

mean that alleles a and A of a gene tend to be reproduced in ratio $(1 - s):1$ per generation. Then the distribution $(1 - q)a + qA$ becomes

$$\frac{(1 - s)(1 - q)a + qA}{1 - s(1 - q)}.$$

The change of frequency of A is the contribution to q by selection pressure. It is given by

$$\frac{q}{1 - s + sq} - q = \frac{sq(1 - q)}{1 - s(1 - q)} \cong sq(1 - q) \quad [3]$$

if s is small.

Migration. If a population is not completely isolated from other populations with which it can interbreed, there will be an inflow and a leakage of genes through migration. Migration pressure thus operates in the same way as mutation pressure on a partially isolated population. If the average frequency of a gene in the entire cellulated population (consisting of partially isolated subpopulations) is q_m, the contribution to Δq due to migration will be

$$-m(q - q_m) \quad [4]$$

where m is the excess of inflow over the outflow of the gene in question.

The random drift. The genetic complex is determined to a large extent by the mutation, selection, and migration pressures. However, there enter also random elements. To begin with, the coefficients of the above-mentioned components of systematic drift may themselves undergo random fluctuations, which (except possibly in the case of mutation rates) may be considerable. Another source of random fluctuations is the so-called sampling variance. Each reproducing generation is only a sample of all the individuals born into the entire generation. If the population is large, the successive samples will not show large deviations with respect to their genetic complex; but if the population is small, the random fluctuations of the sample may be considerable. Thus, the components of systematic drift, by determining Δq, and the components of random drift, by determining $\sigma^2_{\delta_q}$ in Equation 1, determine the gene frequency distribution $\varphi(q)$.

Let us now see how the evolutionary fate of a population can be inferred from the gene frequency distribution. For simplicity, we will consider just migration and sampling variance and examine their effects on the variation from generation to generation in the representation q of some gene. Suppose, also for simplicity, that q_m, the average frequency of that gene in the entire population (of which we are considering a subpopulation) is $\frac{1}{2}$. The contribution of migration to Δq is given in Equation 4. The contribution to $\sigma^2_{\delta_q}$ by sampling variance is given by the well-known formula

$$\sigma^2_{\delta_q} = q(1 - q)/2N. \quad [5]$$

Substituting Equations 4 and 5 into 1, using $q_m = \frac{1}{2}$, and simplifying, gives

$$\varphi(q) = \frac{\Gamma(4Nm)}{[\Gamma(2Nm)]^2} q^{2Nm-1}(1 - q)^{2Nm-1}. \quad [6]$$

Here Γ is the well known "gamma function" defined by

$$\Gamma(x) = \int_0^\infty t^{(x-1)} e^{-t}\, dt.$$

Curves for various values of Nm are shown in Figure 1 (cf. Wright 1932). Note that for both m and N large, the spread of the genetic distribution is small, and vice versa. Other things being equal, therefore, a large migration pressure and a large population favor small fluctuations in the representation of a gene, hence a small drift effect; while a small migration pressure (isolation) and a small population favor a large drift effect. Similar conclusions hold for the effects of selection pressure, and analysis has covered the combined effects of selection, migration, and population size.

The landscape of adaptation. A genotype can be viewed as possessing an over-all degree of adaptation to its environment, a magnitude represented by a real number. The genotype itself, having thousands of components, cannot be represented by a real number but only by a vector having thousands of dimensions. Thus the domain of the variable representing the genotype is a space of thousands of dimensions. For qualitative discussion of the situation, however, consider

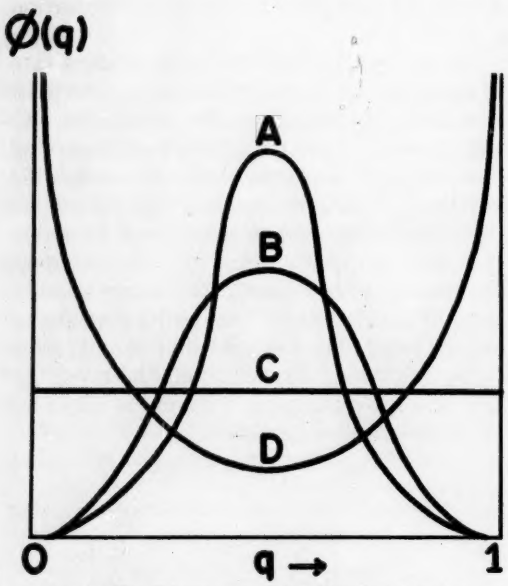


FIG. 1. The Frequency Distribution Curves of a Gene for Various Values of Nm
A: $Nm \gg \frac{1}{2}$; B: $Nm > \frac{1}{2}$; C: $Nm = \frac{1}{2}$; D: $Nm < \frac{1}{2}$

this space as having only two dimensions, the smallest number which can be associated with a "region," and the population as occupying such a region in the genotype space. The area of the region is the greater, the greater the genetic diversity of the population. Moreover, because of the drift effect, the population wanders more or less rapidly over the region, depending on the strength of the drift.

If we think of the degree of adaptation as a third dimension, our two-dimensional genetic complex space becomes a "landscape," in which the regions of high adaptation are peaks and those of low adaptation, valleys. Natural selection of the genotypes of high adaptation guides the population to the peaks of the landscape, but the random drift enables the population to leave them. Moreover, because of changes in environmental conditions, the landscape itself is shifting, so that peaks may change into valleys, and vice versa. The evolutionary fate of the population is now seen to depend: first, on the ability of the population to find an adaptive peak; second, on its ability to cling to its peak; third, on its ability to readapt as the landscape shifts because of changes in environmental conditions; and finally, on ability to find other, perhaps higher, peaks of adaptation even if the environment remains constant.

Mathematical considerations, similar to the above, of the effects on the genetic distribution of the various evolutionary factors and of population size lead to the following conclusions:

1. Under stable environmental conditions, a large, randomly mating population (i.e., where the migration pressure is so large that the entire population can be said to interbreed freely) will remain situated around the particular relative maximum of adaptation it happens to have reached, its "niche." If selection pressure decreases, say under a milder environment, or mutation pressure increases, or both, the area occupied will increase; if converse changes occur, it will shrink; but the population will not wander away from its peak.

2. Under changing environmental conditions (where the relative survival advantages of the several genes change and hence where there is a shift in the peaks and valleys of the "adaptation landscape"), a large, randomly mating population will shift its position *adaptively*, i.e., to the nearest available new peak. Given sufficient heterogeneity (insured by a large population and random mating), the chances of finding a new peak are good; hence there is insurance against extinction.

These two results can be summarized as follows: A large, randomly mating population, having reached an adaptive peak, tends to remain there under stable environmental conditions. It is a safe but "conservative" population. It stands a good chance of surviving, but will not, in general, evolve by finding distant, possibly higher, peaks of adaptation.

3. A small homogeneous population occupies a small area on the adaptive landscape. Under changing environmental conditions, it may find itself entirely in a valley; that is, without a pool of genes adapted to new conditions, and so can easily become extinct. Moreover, such a population is subject to considerable random drift, and so will wander over the landscape, even without en-

vironmental change. Hence, even under constant conditions, a small strongly inbred population may, in fact must, eventually become extinct. Such populations are "adventuresome," they wander; but they are not safe, they are more likely to end in a valley than to find another peak.

4. A "cellulated" population, consisting of several relatively small and relatively weakly interbreeding subpopulations, occupies several small areas of the landscape. The subgroups, because of their smallness and relative homogeneity, do wander over the landscape, even in the absence of environmental changes; but, being interconnected by migration and interbreeding, the total group enjoys the safety of a large heterogeneous population. If one such subgroup wanders into a valley, it can be "rescued" by the others, since the good genes of the other subgroups are available to it. Moreover, such a population is not hemmed in by the conservatism of the randomly mating population; for subgroups are able to wander and explore the landscape for new peaks, while the interbreeding insures the entire population against extinction. Such a population is thus both "adventuresome" and safe; as emphasized by Sewall Wright, it is endowed with the greatest evolutionary potential.

LINGUISTIC EVOLUTION

If by evolution we mean the cumulative process of small changes, it is a truism that this term can be applied not only to the history of organisms but equally to many other histories of development. One speaks of evolving institutions, evolving languages, evolving ways of thought. Often the history of an artifact presents a striking similarity to that of an organism: witness the changing appearance of the automobile, in which we can trace "vestigial parts," such as the whip stands in the earliest models and the gradually disappearing running boards in the later ones.

In the realm of culture we decided to examine phonetic change, partly because of precision and some good documentation in this field, partly because phonological evolution best parallels the "unconscious,"[5] "automatic," and "spontaneous" aspects of biological evolution; partly because language is the best approximation of pure culture; but most of all because, of all aspects of the study of culture, linguistics has not only the most distinctively cultural methods but also the best defined concepts. Linguistics has, mainly within the past generation, grown out of philology. Philology is a humanistic discipline with an interest in endless particulars, of and for themselves; linguistics is a science interested, like all other sciences, in discoverable regularities. The discovery of such regularities presupposes standard fundamental entities like the chromosome and the gene. There are such in linguistics and especially in phonology.

Any humanly produced sound is a "phone." The number of phones is limited only by the anatomy and physiology of the human speech organs as these operate in varying environmental situations. But every language has a limited number (the known range at present is between 11 and 67) of distinctive sound-classes known as "phonemes." Most phonemes have one or more alternative forms (more often than not positional), which are called "allophones." A brief sketch of phonemic theory is now essential to our purposes.

The human vocal apparatus allows the making of a great variety of sounds. But children in each different group soon learn to make, with small variations, a particular set of sounds that has been selected by their culture out of a vast number of biological possibilities. Every sound system has a number of structure points, the phonemes. Each language, in other words, consists of discrete units of short duration which meet certain physical requirements. A phoneme is variously defined, technically, as "a minimum same of vocal features" or as "the class of all segments and spans containing a given feature." The meaning of these (to the nonlinguist) cryptic utterances will become clear if

[5] Edward Sapir speaks of language as "the mountainous and anonymous work of unconscious generations." The facts that (*a*) most phonetic change is little complicated by purposive and value factors and (*b*) that it is massive in comparatively short time periods, give phonological evolution at least certain superficial similarities to biological evolution.

we turn to concrete illustrations from familiar material.

Take the English word "pin." There are three irreplaceable sound units here. Each occurs in other combinations: the first occurs, for instance, in pip, pill, and pocket; the second in pip and pill but not in pocket; the third in tan, run, and hen. But these three units cannot be analyzed by further partial resemblances. In the case of "pin," the three phonemes are represented by three letters of our alphabet. Our conventions of writing, however, are far from a trustworthy guide. In "thick" the first phoneme is represented by two letters, "th," and the third phoneme by two more letters, "ck." The linguist represents the first of these phonemes by a Greek letter (theta), the second by simple "k."

Phonemes are types or classes of sounds which are treated by speakers of that language as units. Thus the "p" sounds in the words *p*in, s*p*in, and ni*p*, which we English speakers react to as if they were the same, are, from the standpoint of an acoustic engineer, quite different. But they are all *allophones* of the same phoneme, "p." In English, "t" has five allophones. The English phonemes "p" and "t" are conventional or cultural, not natural, categories. Nature, as reflected in the sound waves recorded with oscilloscopes, reveals the three distinct sounds for "p" indicated above and the five for "t." But, where the natural world presents nothing but an indefinite number of contingent varieties, the intervention of culture creates a bundle of *distinctive features* that have a minimum sameness and are treated as a unit. In one language, a sound made with the two lips, voiceless, strongly articulated, and with stopped breath is bundled together and contrasted with another sound that is otherwise identical but made with vibrations of the vocal chords (voiced); while in another language, sounds are contrasted on the basis of force of articulation rather than of voiced-voiceless. Lip rounding in the production of vowels occurs in many languages; but it is phonemic in some (e.g., German and Hopi), nonphonemic in others (e.g., English and Shawnee).

Phonology, with its demonstration of ultimate significant entities (phonemes and their distinctive features), thus converges with modern physics, which has revealed the granular structure of matter and the elementary particles which compose it, as well as with biology, with its particulate genes and chromosomes. The number of contrasts (binary oppositions; compare these with the diploid state) of distinctive features in the languages studied this far runs approximately from six to twelve. (This includes only "inherent" features; "prosodic" features would add another three or four.) In the French phonemic system (13), for instance, the indispensable binary oppositions are:

vowel-consonant
nasal-oral
saturated-diluted
grave-acute
tense-lax
continuous-intercepted

Sound change. Although sounds and phonemes are rather more characterized by inertia than the rest of culture, there is constant variation in detail just as there is in all biological behavior. No two speakers of the same language or dialect utter "the same" sounds in precisely identical ways. Indeed, the self-same speakers vary in their production of a phoneme in fixed position, depending in part on the social situation and on their biological state at the moment of utterance. Not all of these variations, of course, can be compared to mutations. Many are perhaps crudely comparable to nontransmitted, somatic mutations or to modifications in gene effects induced by the environment; but some sound changes, like the germinal cell mutations which do spread by inheritance, may become durably established in a population. In fact the striking degree of established regularity requires emphasis here. Grimm's Laws on sound shifts in Western European languages were the first proof of statistical uniformities in human behavior. Since then there have been massive documentations of orderliness in linguistic change. Two illustrations must suffice. In Latin, between 450 and 350 B.C. the consonant "s" was changed to "r" wherever it came between two vowels—*genesos*,

the possessive singular of *genus*, became *generis*; *meliosem* ("better") became *meliorem*. In Hausa, an African language, during the eighteenth century "s" changed to "sh" whenever followed by one of four particular vowels. Much of the regularity in sound change is due precisely to the fact that we have to do with phonemic *systems*. If one phoneme alters radically, there will be compensatory alterations in others, particularly those made in the same region of the mouth or produced by similar mechanisms. If, for instance, a new "back" phoneme "crowds" the "back" series, then a germ of instability is introduced into the whole system. And the introduction of a new principle ("distinctive feature") or the dropping out of an old one will gradually produce changes spreading across all the relevant areas of the sound system.

There are, however, regularities other than those deriving from the systematic nature of phonemics. Some of these are analogous to the ecological factors, such as location, radiation, and the like, that have proven so significant in biology. It has been shown (2), for example, that the following principles generally hold:

1. Isolated areas (such as islands[6] or mountainous regions) conserve older linguistic features than others.

2. Lateral areas preserve older linguistic features than central areas.

3. Of two forms, the older one is spread over the larger, the innovation over the lesser, area; provided the latter area is not either the isolated one (norm 1) or the sum of both lateral areas (norm 2).

4. Of two forms, the older one is preserved in a territory later occupied (*area seriore*) or colonial. . . .

Some analogies. Though we shall make appropriate and necessary qualifications at points, our purpose, in general, is to press the biologic-linguistic analogies as far as seems possible, perhaps further. Note the following, as a beginning:

1. There appears to be almost as much order and regularity in linguistic evolution as in biological evolution.

2. In both cases, "internal necessity" (genetic and linguistic systematic "drift"); selection to meet adjustive and adaptive needs; and external pressures (migration, location, contact) seem to be supremely relevant.

3. The two sets of phenomena reveal similarities due to "convergence" as well as to genetic descent. The great linguist, Troubetskoy, contrasts language groups that are analogically similar (*Sprachbund*) with those that are homologous (*Sprachfamilie*). The existence of similar forces (internal and external), operative under similar conditions, is indicated by the existence of a limited number of types of changes which reoccur in historically unconnected languages and at different chronological periods.

4. In both phonology and genetics there are "accidental" (i.e., not predictable within the limits of the theory) abrupt changes. Geneticists cannot yet explain why "spontaneous" mutations occur, historical linguists cannot yet say why accent suddenly took the place of vowel quantity in a sound system.

5. In both realms one operates with a limited number of fixed loci. Just as every species is characterized by a definite number of chromosomes, so every phonological system at any given point in time is organized in terms of a specific number of phonemes. One can speak of loci in both instances. Some languages have more loci than others, but all languages have vowels, semi-vowels, consonants, at least one nasal and sibilant and at least two stops. There is a restricted number of articulatory positions (e.g. the tongue can be alveolar, dental, cacuminal, retroflex, kettle-shaped, etc.) and of sound types (stops, fricatives, affricates, clicks, and the like).

6. There appear to be linguistic analogues to the "position effects" on genes; most allophones are positional variants. Every language is made up of a combination of sounds, the vast majority of which have no meaning taken by themselves. The environment of the individual sound is determina-

[6] An island is, of course, not necessarily "isolated" in the sense meant here—e.g., England for many centuries.

tive. Similarly, no or few genes will produce identical effects regardless of the contribution of other genes which may or may not be present in particular positions. This shades over into linguistic analogues to modifier genes. Thus, in Navaho, only the presence of a long vowel or diphthong in a word makes possible rising or falling tone, as opposed to low and high tone. The high-low variation is possible alike with short vowels, long vowels, and diphthongs.

7. There are arresting, though perplexing, analogues between the fundamental entities themselves. In some ways the most satisfactory parallel is between chromosomes and phonemes, genes and distinctive features, and allelomorphs and allophones[7]; we have, however, not yet worked this out in an altogether acceptable way. At least, as chromosomes are made up of bundles of genes, so phonemes are resolvable into distinctive features. A listing of the relevant distinctive features constitutes, in fact, a definition of that phoneme. No distinctive features can be further decomposed by linguistic analysis.

We recognize fully that certain combinatorial principles in genetics do not apply in linguistics. This is most notable as regards the biparental source of genes in organisms reproducing sexually, with each parent contributing only the haploid number of chromosomes. It is also true, of course, that most of an individual's genes remain unaltered through life, whereas many of the phonemes of both individuals and groups can change within a generation. Nevertheless, let us boldly explore the analogies that appear in the four rubrics used in population genetics.

Mutation.[8] There are sudden sound changes, both regular and, sometimes, sporadic. An example of a sporadic change is the replacements of Latin "r" by "l" in the Spanish word for "tree" (arbol). Some examples of regular changes are: fricatives in proto-Indo-European became stops in Proto-Germanic; in early Aramaic a voiced labial spirant was replaced in all instances by a voiced labial stop; Vulgar Latin (and the Romance languages) replaced vowel quantity by stress accent. Most often, perhaps, mutations begin with allophones, leading later in some cases to new phonemes.[9]

Selection. It was a dictum of the older linguistics that "sound change is blind." In Old English, intervocalic "s" as in *céosan* shifted to "z," whereas "s" remained in other positions. But Old Icelandic preserves a voiceless "s" in *kjosá*, and many other languages show no sign of voicing their intervocalic sibilants. Linguists were fond of examples of this kind to back up sweeping statements that the causes of sound change were unknown. Today linguists would modify this generalization to

[7] The first two of these analogies derive from viewing the phoneme as a set of distinctive features, each of which represents a selection from a binary juxtaposition. If one continues this analogy, however, one arrives at the juxtaposed pairs of the distinctive features, rather than the allophones as analogues to the allelomorphs. As indicated above, the allophones are more reminiscent of the position effect of genes. Still, in another way, the allelomorph-allophone anology is suggested in the process of phoneme modification in which one allophone changes into another until the phoneme is replaced altogether. One might also emphasize the highly functional significance of allelomorphs by comparing them to phoneme alternants. Thus, at one locus one may get a gene for antigens A or B or the gene for the absence of both (Blood Group O). But one necessarily gets *one* of this series (including alternant forms for antigen A). Similarly, at the x "stop locus" in language one may get at bilabial positions one of the following voiceless forms: p, preaspirated p, postaspirated p, or voiceless lenis p.

[8] The documentation in this section of the paper is drawn mainly—though not exclusively—from the articles by Martinet (18) and Greenberg (8). Many of our statements are taken directly or in modified or paraphrased form from these publications. The original sources should, of course, be checked on any point deemed important. But, since we are not writing a contribution to technical linguistics, it would be pedantic to cite page numbers and modifications in detail. On the other hand, since more standard modes of biological analysis are here applied in a more novel manner to language, we have presented more sources and annotations in this section.

[9] Is there a parallel here to genetics? Bentley Glass (7) speaks of, "The origin of new genetic loci from old ones by the occurrence of small duplications or 'repeats', in the chromosomes, followed by a gradual differentiation of function and increasing independence of action, until the new loci have passed beyond the stage of pseudoalleles into that of clearly distinct genes."

the effect that some of the causes are either unknown or difficult to identify and verify. More important, much of what once appeared as "blind change" now submits to analysis in terms of phonemic theory. Some changes affect phones in all their contexts; others only phones in phonemically well-defined contexts.

Selection certainly operates in linguistic evolution in at least two respects: (a) in terms of sheer human laziness—Zipf's (30) "principle of least effort"[10]; (*b*) in terms of communicative needs of speakers. Some single sounds are relatively rare in the languages of the world because they are intrinsically "difficult" to make, others not so much because they take more work or effort but because they have to be made just right. Unaspirated voiceless stops need critical timing; fronted palatal stops need critical pressure of air. Certain combinations of features and sequences of sounds present articulatory difficulties, and hence they tend to be smoothed out or eliminated. For example, a tendency toward palatization of front vowels has been observed in hundreds of languages. This simply makes these sequences easier to say. Such progressive changes as that of "Caesar" to "César" are dependent on a shift from the slowly-moving rear tongue to the faster moving tongue tip, in articulating consonants. The liaison and the dropping of final consonants in French are due to the increase of rate with light stress, which forces the arresting consonant to shift to the next syllable in the syllable train or to disappear (22, p. 98).

No one would claim that changes in morphology, syntax, and lexicon are uninfluenced by communication problems; yet the significance of the sender-receiver relationship is equally clear for phonology. Phonemic change involves, in decreasing order of frequency:

a. replacement of one phoneme by another (much the most common)
b. loss of a phoneme
c. transposition of phonemes
d. insertion of a phoneme

If the distinctive features, governing significant contrasts between two phonemes of a group of phonemes, alter, sounds that did not previously occur in the language may appear or a new patterning of existent sounds may result. In the latter case there often occurs the phenomenon of merger, involving a reduction in phonemic inventory, and the regularities on merging[11] illustrate both the principles of least effort and of communicative need:

1. The more uncommon a phoneme is in human speech in general, the more likely it is to be merged with another phoneme.

2. The lower the frequency of a phoneme in a given language, the more likely it is to merge with another phoneme, providing this second phoneme is not itself of excessively high frequency.

3. The closer the points of articulation shared by two phonemes, the more likely they are to merge.

4. The more distinctive features shared by two phonemes, the more likely they are to merge.

5. The fewer the pairs of different linguistic forms which are distinguished by two phonemes, the more likely they are to merge.[12]

The communication angle also comes out clearly in some recent collaborative work between linguists and communication engineers (8, pp. 156 ff.; 17). The number of distinctive features needed to describe the

[10] However, Zipf's thesis that "difficult" sounds are less frequent than "easy" sounds in the languages of the world needs further definition of conditions and in any case appears to be a great deal less than an exceptionless uniformity. It probably also needs to be supplemented by a principle of probability to the effect that "events corresponding to a greater range of combinations of factors are more likely to happen than those limited to a narrower combination of factors." See the review by Chao (4).

[11] Mergings are often "avoided" by compensatory shifts throughout the system. See especially Martinet (18). This appears to be partly a matter of acoustic and articulatory economy, partly a matter of preserving the symmetry of the system (see below under "Drift.")

[12] For example, in English the functional yield of the "p-b" opposition is very high. Hundreds of frequently used words are distinguished by this contrast. On the other hand, only a few pairs are contrasted by another voiceless-voiced opposition: thigh-thy; mouth (noun)-mouth (verb).

number of phonemes would approach the minimum of $\log_2 n$, where n is the number of phonemes in the system. But since the communication needs of both speakers and hearers must be taken into account, it was argued on theoretical grounds that factors of efficiency and redundancy should be equal.[13] The actual number of distinctive features thus is predicted to be close to double the minimum number, permitting a redundancy of 50%. In languages which have been analyzed this estimate has been borne out. In contemporary American the figure is 53%; in contemporary Russian, 48.9%; and in Spanish, where records at four time-points have been analyzed, the number oscillates over this mean—48.8, 57.3, 47.5, and, today, 50.0. In other words, the equilibrium between efficiency and redundancy[14] is maintained or, if the system gets out of line, rather quickly restored. Inasmuch as sounds are put together, among other things, to compromise between the needs of the speaker and the listener, it follows that the evolution of phonologies is subject to functional selection, and is not merely "accidental" and "spontaneous." Linguistic evolution in general can be conceived as regulated by the fluctuating balance between the expressive needs of man and his propensity to keep his physical and mental exertions at a minimum.

Migration. Just as genes "flow" from population to population, so do sounds—a subject of great interest to Franz Boas (1). Some Bantu-speaking groups borrowed clicks from Bushmen and Hottentots. Some Indo-European languages of India took over cacuminals from Dravidian or other pre-Indo-European tongues. Centuries after contact between peoples of different languages has ceased their languages may "perpetuate evidence" of the earlier contact (29). The problems of bilingualism (24), generational differences (21) and the like are excessively intricate. Casagrande (3) says:

> Linguistic change may occur in all the various aspects of language: phonology, grammar, vocabulary, and style. However, modifications in the highly systematic aspects, phonology and grammar, appear to progress at a slower rate, and while changes in these realms of language may be accelerated or even initiated by culture contact, they seem to result primarily from relatively autonomous linguistic drifts.

Bilinguals have, so to speak, two "phonological personalities." They speak language A with the "prejudices" of language B and vice versa. When bilinguals become imitated in respect of one or more features by the bulk of the speakers in a speech community, then a change becomes established.[15] A kind

[13] This hypothesis receives some confirmation from the fact that the lowest frequencies of phoneme clustering are between phonemes maximally similar or maximally different. The results for 845 consonant clusters in a series of 20,000 phonemes tend to follow a normal curve. "It seems justified, then, to assume that at least in consonant clusters, maximum efforts for either encoder or decoder are avoided in favor of those situations where the effort is more or less equally divided" (8, p. 102).

[14] A somewhat forced analogue to "redundancy" in genetics: in *Drosophila* there are many different genes (far more than are "needed") which will produce eye color. Presumably this is "multiple security" for developmental processes.

[15] It may help to illustrate this very concretely from a known bit of linguistic history:

> In Old English there was a phoneme /f/ which had the allophone [f] initially, when doubled internally, and finally, and the allophone [v] when single internally ('five' had initial and final f, *seofen* 'seven' had internal [v]). This was part of the formal system of the old English language. After the Norman conquest . . . , words from French began to be used, informally by individuals who 'picked them up,'. . . . In Old French /f/ and /v/ were separate phonemes (*femme* but *vain*, *affaire* but *avoir*— and only /f/ in final position, *vif*). The English speaker who picked up such an old French word as *veal* 'veal, calf' (modern *veau*) could either pronounce it with initial [f] /f/, following his formal system, or could use initial [v] as an informal deviation from that system. When, however, French words became more widespread, the use of [v] in such instances became technically demanded and approved; we can imagine the kitchenhand from his lord's castle telling his friends back home about how calf meat is called *veal*, and when some one of them tried to say it and uttered an initial f, correcting him explicitly, 'No, not [f . . .], but [v . . .].' Finally, the technical knowledge of how to use [v] initially became widespread, the word *veal* (and others like it) became a part of English, and a new formal system was thus instituted, in which /f/ and /v/ became separate phonemes, as they continue to be. When the old formal system was replaced by the new one, the previous formal usage became an informal one—some speakers still said [f] in *veal*,

of contagion takes place. Swadesh (23) generalizes some of the processes as follows:

> The principle that seems to govern language replacement, both with respect to dominating and dominated groups may be stated thus: In an area and era of broadening economic and social intercourse, there tends to be more and more widespread use of fewer and fewer languages. Speakers of more restricted languages tend to become bilinguals and eventually, often after a number of centuries, the minor languages are eliminated. Factors in language rivalry include: population numbers, breadth of geographic spread, mobility of the population, economic and political dominance and activity. Numbers are a highly important factor but may to some extent be offset by other factors. In places and times (e.g. medieval Europe) of reduced intercommunication, linguistic communities suffer gradual fragmentation with local dialects of a single original language developing into separate and mutually unintelligible languages (23, p. 344).

Weinreich (24) finds four types of phonological interference: three phonemic (underdifferentiation, overdifferentiation, and reinterpretation) and one phonetic (phone substitution). Even phonemes that already occur in a language become difficult when introduced in new sequences from a contact language.

Drift. There is little doubt that some features of phonological evolution are determined by tendencies immanent in the sound system at some earlier time-point—a type of systematic drift.[16] This is shown by the fact that many changes in sundered speech communities follow remarkably parallel lines, even though the separate groups move into varied environments and are exposed to different historical and biological vicissitudes.

Trends more dependent on the internal dynamics of the system than on external selective factors (toward "function" as regards economy and communication) are also illustrated by the tendency to preserve symmetry in phonemic pattern. If, for instance, a language borrows through foreign contact ("migration") a voiceless velar spirant, it is probable that this will eventually be "paired" by the adoption of a voiced velar spirant, provided, the voiceless-voiced binary opposition is elsewhere prevalent in the system. There is no functional "need" for this. It is just a question of filling in what French and Spanish linguists call "the empty boxes"—i.e., maintaining neat symmetry. This is probably why such phonemes as ž and š do not approach each other (tending to merge) in English, in spite of the exceptionally low yield of this opposition. Though this specific form of the voiced-voiceless contrast is functionally insignificant, speakers (unconsciously) insist on maintaining it because it is an instance of a more generalized principle.

In other words, phonological evolution is seldom discrete. Because phonemics are involved, the evolution of sounds is always complicated by the pattern relations and hence cannot be explained altogether in functional ("selection") terms. If a "t" in the system "mutates" to aspirated "t" through unknown causes ("historical accident"), other voiceless stops in the series will sooner or later get aspirated. There appears to be no "functional" sense in this process; the "internal logic" of the system ("the theory of pattern attraction") just demands that all phonemes of a pattern become as fully symmetrical ("integrated") as conflicting factors make possible.[17]

A firmer analogue to drift in the genetic sense is arguable. We recall that in the evolution of biological populations the random drift effect is strongest among small, highly inbred populations. Whether or not a similar effect in linguistic evolution is a consequence of small size and inbreeding requires more research than we have been able to accomplish; but there are fragments of evidence to support the hypothesis. Languages having numbers of phonemes

but this was no longer 'correct' or 'standard' or 'proper' (9, p. 31).

[16] For a brief summary of "linguistic drift" see Harris (10).

[17] "Symmetry" or "internal logic", despite many violations, applies to grammar and other aspects of language as well as to phonemes; "mouses," and, "deers," are likely to be used by young children. Indeed, such expectation of regularity is built into all nervous systems; as shown by the conditioning and generalizing behaviors, which underlie induction, as well as by the splitting of experience into categories, which is basic to deduction.

within the known minimal range (Polynesians, Arunta, etc.) happen to be spoken by peoples who lived for many centuries in relative isolation from languages of markedly different sound types. Conversely, phonemic differentiation is most marked among small groups living where the communication networks were most complex (12).

The linguistic differentiation of the small groups along a relatively short but rugged coastline, as in Oregon, indicates the probable importance of the terrain. Here rugged mountains running to the coast separate the valleys of short rivers which flow from the Coast range; and linguistic (and probably population) differentiation is found between valleys. Moreover, the phonemic richness of speech is greatest at the valley heads, where cross valley traffic is greatest. The Caucasus, likewise an area where phonemic systems of strikingly different sorts were in sustained contact, is characterized by numbers of phonemes approaching the known maximum. The aboriginal New World was presumably populated by successive waves of migration from Asia, so the Americas had a plethora of families of unrelated (or possibly, in some cases, very dimly related) languages. American Indian tongues, in general, have a number of phonemes that is higher than that in many areas of the world. We hazard the speculation that a "function" of phonemic elaboration may be that of accentuating and preserving general cultural differentiation. Certainly the Caucasus and aboriginal America north of the Rio Grande represent locales where there is a high degree of "cellulation," the mixture of isolation and intercourse.

Here seems to be an analogue to the cellulation of partially isolated but interbreeding populations. It should be noted, however, that the principal evolutionary effect of cellulation in biological populations concerns their "evolutionary potential," i.e., the combination of variability and safety from extinction, and the analogue of this effect in the linguistic situation is not immediately clear. Partial isolation of small linguistic units and lively interchange with contact languages seem to be associated with phonemic richness, in contrast to the phonemic poverty of both the totally isolated units and (to a lesser extent) large thoroughly united ones; but it is not clear just how this phonemic richness affects either the viability or the rate of change of a language.

Although the complexity of the cultural-historical processes may preclude a simple formulation of equilibrium, such as the Hardy-Weinberg law of gene distribution, nevertheless effective equilibria are maintained—as evidenced by the findings on efficiency and redundancy. Perhaps, in principle, equilibrium is as important in the linguistic as in the genetic case but is not so finely demonstrable in detail, because historical accidents do not cancel each other out as completely in the cultural realm as in the biological. On the other hand, one must not demolish the various parallels too cavalierly, either. Thus, it is easy to say that, while alleles are interchangeable, phonemes are not; but the facts are (*a*), that some phonemes are interchangeable, and (*b*), if gene function is considered, alleles are *not* altogether interchangeable.

It is certain that the spoken language of the individual is culturally inherited—predominantly from parents, parent surrogates, and siblings. The "inheritance," however, is not a simple duplication but a learning process, involving trial and error, social reward and punishment, with a trend toward a pronunciation most closely resembling that of preferred elder people, including siblings. Moreover, this learning process need not be confined to the period of initial speech formation but may continue for years, so that a person's speech may change several times as his important daily contacts change. Thus, an individual's linguistic "phenotype" is not determined by factors inherited from two parents and combined by fertilization, but by a more complicated process, involving often a great many "parents" and a gradual stabilization of the phenotype in time. We might, to be sure, view the modification of the individual's speech as constituting a series of "somatic" mutations. In biological systems, however, somatic mutations do not significantly affect the fate of the germ plasm, which is the carrier of inheritance in

organisms possessing sex; and there seems to be no equivalent of the "germ plasm" in language. The analogy may be closer, however, with parthenogenetic species in which biparental inheritance is absent and plasma genes are more important. Even in protozoa, with endomixis, plasma entities, as the killer factor of paramecium, do play an important role. The recently discovered ability of viruses to carry genic material from one type of microorganism to another—transduction—is also relevant. Evolution is thus possible by a series of small changes during the life span of individuals, not unlike learning. The best analogue of the genotype, however, is the inferred or constructed "ideal" structure of the total phonemic system.

Several possibilities present themselves as the analogues of the organism: the single instance of the spoken word, the word as it is ordinarily spoken by a particular speaker, or the total verbal output of a particular speaker. If the first is chosen, the "organism" is ephemeral, indeed; but it has a large succession of "descendants," the successive instances of the same word spoken by the same speaker. The progeny goes through an "adaptive process," as the pronunciation of the word becomes stabilized in the early years of the speaker; but it may also undergo more discrete "mutations," as the pronunciation is modified by other factors. If the word as most frequently pronounced by a particular speaker is to be taken as the "organism," then we should consider the stabilization of phones as the ontogenetic development of the organism, while the modifications of speech in successive generations of speakers would be viewed as the phylogenetic development. If the organism is taken as the total verbal output of the individual, the considerations just above apply, except that the entire phonemic repertoire rather than the word is to be viewed as the "phenotype."

In any case, "somatic mutations" or "modifications" are transmitted in language, if by transmission we mean the modification of the speech of one individual by that of another. If the single instance of the spoken word is taken as the "organism," the above difficulties are eliminated—but others take their place. Shall we take the totality of spoken instances of the same word or the total phonology as the "species"? Presumably the latter. If the former, then the interaction of "species" becomes much more intimate than in biology, so that one cannot consider a species in relative isolation, as is done as a first approximation in genetic theory. The laws of "species" interaction probably acquire, under this assumption, a crucial importance that must somehow be discerned. If the total phonemic system is taken as the "species," then the "species" consists of thousands of grossly dissimilar phones ("individuals"). These are, however, strongly interdependent and a generalized phone can, to a considerable degree, be identified by phonemic theory. As stated, the "ideal" structure of the phonology attributable to a linguistic community resembles a "genotype"; but if one looks primarily at the enormous concrete variability of phones, language appears less like a species and more like an organism, consisting of interrelated tissues and organs that vary through time and develop more by an embryological than a genetic process.

"Mutation pressure" in language, the tendency to develop "new" allophones or phonemes, instead of remaining relatively constant for particular units may well be dependent on a great variety of factors. (In genetic theory, mutation pressure is taken as relatively constant for each gene, but it does probably vary over a longer time scale; indeed, a number of the problems of analogizing here stem from the different times involved in biologic and social evolution.) "Mutation" would not include changes in phones brought about by bilingualism or sustained contacts with "foreign" speakers—these would be subsumed under the factors of selection and migration—but refers to spontaneous, nonimitative linguistic changes. Some of these are occasioned by influences other than the speech of other individuals, though the speech of others may often be a stimulus or "catalyst." Situational changes (climate, nutrition, flow of different biological genes into the

population) can bring about alterations in modal "constitution" or "temperament" and these, in turn, are reflected in variations in standard respiratory patterns, speech tempo, stress, and the like. Other innovations in phonology may not indicate anything more than that propensity for variation which inheres in living matter and which becomes expressed in behavioral variation. But even in these "genuine" mutations (as distinguished from imitative-adaptive modifications), the direction of change may be influenced by environmental conditions—a circumstance certainly unproven in biology.

In considering "selection pressure," we must again discriminate between selection as it operates in bringing the speech of one individual nearer to that of another (by reinforcing certain responses and inhibiting others) from the selection pressure which is imposed by other, nonimitative factors. Under the latter type of selection pressure are those factors which favor certain sounds and sound combinations for anatomical and physiological reasons of "least effort." In general, innovations can be expected to be selected which interfere least with established motor habits of speaking.

At any rate, a definition of survival advantage of a speech unit analogous to the survival advantage of a gene could be constructed positively in terms of the potential spread of the unit through the population (its "fertility rate") and negatively in terms of its tendency to fail to establish itself or to be eliminated (its "mortality rate"). This definition is especially natural if the single instance of the spoken word is taken as the individual organism, and its repetitions (with or without modifications) as its progeny. It is perhaps noteworthy that the modification of the "progeny" may be of two kinds: (*a*) the repetitions of the same word by the same speaker in the absence of instances of pronunciation by others involve "internal" modification based, perhaps, on the needs of the speaker; (*b*) the repetitions following instances of pronunciation of others involve "external" modifications, which, if this particular model is pursued, seem analogous to an exchange of "genetic material."

One important question to be considered with reference to selection is on what units selection operates. It certainly operates on several. Under some circumstances (e.g., cliques, social classes, etc.) comparatively small phonological variations have "survival value." Selection also operates on sound clusters, words, intonation patterns for whole utterances; and in some cases selection may conceivably operate on the speakers. The difference between "s" and "sh" selected for slaughter, at the passages of the Jordan by the Gileadites, all who failed to say "shibboleth."

Finally, we explicitly note an additional complication: phonological change is sometimes initiated and influenced by alterations in the nonlinguistic aspects of the culture—and it is well known that a large number of variables enters into the processes of culture change in general. To be sure, genetic processes are not immune to culture change either—a radically different mating system in a small population will, over the generations, alter the genotype of the population—but some facts on phonetic and culture change, well summarized by Hoijer (11), are impressive.

Lexical borrowings, based on familiarity with new objects and ideas provide new phonetic contexts. For example, two phonemes that never earlier appeared in initial position in Chiricahua Apache began to do so as a result of lexical borrowing from Spanish. This altered, albeit in a minor way, the whole phonetic equilibrium of Chiricahua utterances. (Perhaps the analogue here is to modifier genes.) Or, newly formed compounds may produce sounds that have not previously existed in the language or may bring together phonemes or allophones that have not previously occurred in sequence. Such a disturbance of the phonetic equilibrium may bring in its train a whole series of compensatory shifts which eventually, like the first Germanic consonant shift, can lead to changes in almost every aspect of the phonetic system.

In spite of all the difficulties—and they

are very real, possibly insoluble—we feel that the analogues between biological and linguistic evolution are also real and interesting. Without much straining, one can talk about all of the known major processes of phonological change under the same rubrics (mutation, selection, migration, and drift) that are necessary to the understanding of population genetics. While there are places where the similarities break down and while there are details of significance which do not bridge the two fields, it appears to us that the genetic categories cover all or almost all of the main phenomena of sound change. Moreover, it should never be forgotten that *genetic* categories are also, in some respects, still in flux. Bentley Glass (7, p. 233) has written as follows:

> As to the gene—clearly when we have before us a process of continuous differentiation in the genetic material by means of pseudo-allelism, the nature of the "gene" depends upon the stage in the process. A single biochemical unit, a segment, a repetitive sequence of identical units that can undergo recombination, or a sequence of functionally different pseudoalleles—the gene might be any one of these up to the time when it has become clearly a multiplicity of genes.

The over-all conclusions concerning the evolution and fate of languages should appear, then, as the consequences of the consideration of those aspects which turn out, on sufficient examination, to be analogous to the "genetic distribution." The task of a mathematical theory is to develop an axiomatization concerning the duplication of linguistic "organisms" and their modifications throughout their "lives" and to derive what are essentially equilibrium theorems (analogous to determining the equilibrium representation of a gene subject to mutation, selection, and migration pressures). A mathematical expression derived for the equilibrium genetic distribution of a linguistic population could then lead to predictions concerning the evolutionary fate of such populations.

CULTURAL EVOLUTION—A LABORATORY APPROACH

Although the success of the experimental method in the natural sciences has been so pronounced that its use there requires no further justification, how far this success can be extended to behavioral science is still an open question. Three conditions are crucial: (*a*) the possibility of imposing sufficient controls upon the situations examined so that one can truly talk about "an experiment"; (*b*) the generalization potential of the situations examined; and (*c*) the relevancy of the results to what is held of importance in behavioral science.

A great deal of methodological controversy revolves about the feeling of some social scientists that conditions *a* and *c* are somehow incompatible. With varying degrees of conviction it is asserted that what can be unambiguously (especially quantitatively) described in human behavior is not interesting or important, and that what is important cannot be subjected to rigorous experiment under controlled conditions. The behavioral scientist with an "exact science" orientation, whether he thinks in terms of a controlled experiment or the statistical evaluation of field data or a mathematical model of behavior, probably tends to be attracted toward situations which yield to his methods and tends to attribute importance to them. This frailty, however, has a counterpart—the social scientist who depreciates the exact theoretical, experimental, and statistical approaches may do so because he is repelled by situations which do yield themselves to methods alien to those to which he has become accustomed.

In this section we describe an experimental program designed to simulate the evolution of cultures. Since, as will appear, the situations to be studied do lend themselves to experimentation, conditions *a* and *c* seem satisfied. What of condition *b*—to what extent can the findings in experimental situations legitimately be generalized to the evolution of real cultures in human societies?

The question must be answered by test and it is hoped that two things will be accomplished by pursuing our program. First, techniques will be developed, more appropriate than those available, for experimenting with human beings on levels characteristic of specifically human behavior; levels

involving, beyond responses to stimuli, individual learning, and other traditional matters studied in experimental psychology, such behavior as the evolution of hypotheses and the internalization of unconscious assumptions, their transmission through "generations," the role of tradition, ritual, and communication patterns, the evolution of language itself, etc. Indeed, since the idea or insight is an element that helps generate culture, like a gene, by interacting with its fellows and the environment, and since a new idea is a discontinuous "mutation" of the generating unit, our experimental program, no less than the historical examination of language, is a direct study of analogies between biological and cultural evolution.*

Second, it is hoped that the experimental *in vitro* program will lead to formulation hypotheses concerning the evolution of cultures *in situ*. Not that there is a lack of hypotheses about human behavior, but that new experimental techniques may give rise to a new sort of hypotheses. Insights stemming from controlled experiment have often been more useful than those stemming from intuition or from uncontrolled observation.

Considering, then, the creation of a microculture, the evolution of which under experimentally manipulated conditions could be directly observed, we are faced more with a research program than with a particular set of experiments. The essential invention is to gather a small group of individuals, which is presented with some task to perform under specified conditions of operation, and then to rotate progressively the members of the group. The evolving group "culture" will thus be transmitted to successive generations of individuals, as new members continuously replace old ones. By choice of the experimental conditions, particularly the task, it should be possible to follow the group generation of hypotheses; development of communication, both by spoken words and written symbols; crystallization of rules and customs, roles and statuses; creation of artifacts; or even, perhaps, selection of values. Such a microculture indeed appears in real life and is well exemplified in the individual "personalities" that develop in particular dwelling clusters in planned housing developments and that persist despite rotation of families. Some of the controlling factors, such as the position of the buildings, have been pointed out in the articles on "The Transients" by William Whyte in *Fortune* (25).

Of these possibilities, we have so far explored only the first two—hypothesis formation and communication development. We have considered technical devices which will presumably enable us to manipulate the selection pressures of the environment, the types of "mutation" and the receptivity of the group to them (a kind of internal psychological environment affecting unit flow), the explicit factors of flow and migration, and those of isolation and random drift. In more detail, the experimental technique is described in terms of the variables that can be manipulated and the probable consequences of such manipulation.

A single type of task is shown below in two different examples.

Given the patterns X and Y, construct a pattern in Z on the same principle:

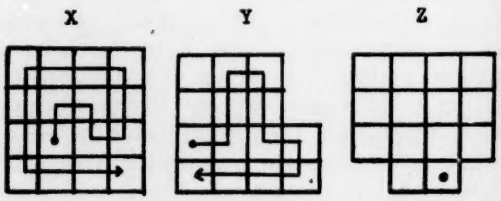


FIG. A

Similarly for X′, Y′, and Z′:

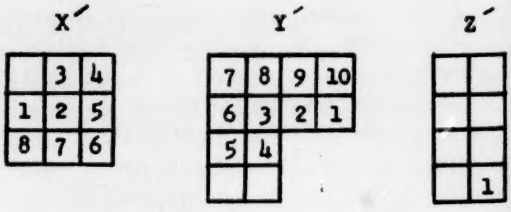


FIG. B

* Added in proof: Our attention has been called to related experiments of E. Rose and W. Felton, "Experimental Histories of Culture," Amer. Sociol. Rev., 1955, 20, 383–92.

The solution is contained in the following rules: The line proceeds to an adjacent square, or integers are placed in succession, in the following order of preference: (*a*) in the direction containing the largest number of empty squares; (*b*) up; (*c*) to the right; (*d*) to the left; (*e*) down. (The shape of the grid and the starting position can be changed arbitrarily.) The solution has not yet occurred to any subject. It is easy to construct tasks of lesser or greater difficulty;[18] indeed, drawing a path through a grid (or even an undivided area) may prove experimentally quite different from placing numbers.

Some basis of scoring the attempted solutions is required. To fix ideas, consider the second form of the task. A score can be devised by comparing the proposed pattern with the correct one (for the same grid and the same position of "1") by counting the mistakes in placing the next number, by counting the number of integers in the proper places, by assigning weights to mistakes (it is a bigger mistake to go down when up is correct than when to the left is correct, etc.), and in various other ways. Consideration of the total experimental situation will be relevant to the method of scoring, as will appear below.

Next, consider a single individual seeking correct patterns. He will learn something and thus improve his score if he gets information about his performance, even though the resulting learning curve will hardly be a consequence of inductive and deductive reasoning but more akin to the internalization of certain apperceptive habits, much as in the acquisition of a muscular skill.

The learning curve can be modified by introducing an additional factor in scoring, which will be of importance in the discussion to follow. A "tradition bias" can be introduced by allowing the "correct" response to move from that initially set by the experimenter in the direction of the actual past performances of the subject. Moreover, the relative weights of the objective score and the "tradition bias" can be varied at will, so that "experimentation" with patterns can be encouraged or discouraged in varying degrees. We note in passing that if the individual develops "theories" concerning the nature of the task, it is possible by means of the "tradition bias" to reinforce any theory he makes: his theories become self-fulfilling assumptions. Another way of discouraging too free "experimentation" is by introducing tolerance limits on the score, below which the subject "dies."

So far we have only a classical individual learning experiment. Our main interest, however, is not in the learning curve but in the psychological and possibly cultural by-products of the situation. The subject will make successive hypotheses about the solution to the problem, the nature of which permits a great number of more or less successful ones. Moreover, the exploration of the consequences of some hypotheses should be impeded by the subject's "commitment" to others. To some the numbered grid may appear as a modification of a magic square problem, so that they seek a rule governing the sums of the numbers in rows, columns, and diagonals. The idea that the rules apply to the ordered placing of numbers may or may not occur. If it does occur, again many hypotheses are possible about where the next number is to be placed—that its position depends on the position of the last, that it depends on the direction from which the last number was approached, etc.

Suppose now that our subject, having experimented up to a point or having "died," is replaced by another, to whom he leaves a "legacy" about "what life is like." This legacy may be transmitted in one or both of two different ways. First, the "tradition" accumulated in the scoring scheme modifies the situation itself. Grooves, so to speak, have been cut in behavior and these are "inherited" by the successor through the modified scoring scheme, even without any communication between individuals. Second, the "ancestor" may actually tell the

[18] For example, this task is practically insoluble: given the series (*a*) 4, 5, 6, 5, 6, 1, 2, 11, 18. . ., and (*b*) 3, 2, 1, 2, 3, 10, 15, 26, 35. . .; continue, by the same rule, series (*c*) 0, 3, 6. . . . The formula is $|2p - n^2|$—the absolute difference between doubles of successive primes and squares of successive integers—(*a*) starting with $n = 0$ (*b*) with $n = 1$, and (*c*) with $n = 2$.

"descendant" what he knows or thinks he knows about the situation. Either a modified environment or symbolic communication may thus serve to transmit acquired habits. In real life, members of one generation may tell those of another how they ought to live, transmitting information through symbolic channels; or they may leave a legacy of organization or technology which itself channelizes life into certain patterns, transmitting "grooves of behavior" without explicit communication. In principle it should be possible to compare quantitatively the relative effects of the two kinds of "time binding" in the experimental situation described.

Consider now a group of, say, three to six individuals engaged cooperatively in the task, but with individual members replaced from time to time by others, neophytes. Our "society" has a birth and a death process and, in due time, all the individuals are replaced. The foregoing considerations still apply, however, and the "culture"— the set of premises, categories, and ideas of how to deal with the environment (the task)—persists, being transmitted either through the internalized tradition bias of the scoring or through verbalized "education" of the newcomers or both. In a group new factors appear, of course, and these seem fundamental to the problem of simulating cultural evolution in a laboratory situation. To begin with, the "education" of the newcomers may be accomplished in at least two different ways, depending on how the task is to be performed by the group. If the efforts of the individuals are independently evaluated and if each works with his own patterns, then this "education" may be simply in the form of advice. If the group score is important, taken as the average of the individual scores, some pressure will be brought to bear to induce conformity to group norms. If, however, all the members of the group must work on the same grid, say, each contributing a number in a square, the pressures can be expected to be more severe on the neophyte, whose lack of experience or excessive enthusiasm for experimentation could lead to the group's extinction through a low score.

By far the most significant group factor expected to arise is intra-group communication. In contrast to intra-individual communication through unspoken thought, the verbal inter-individual communication in the group is amenable to considerable experimental control and also to complete observation, especially if only written communication is allowed. Let us consider this factor in some detail.

The language of group communication can enter as a variable in a variety of ways. In the simplest case, a limited vocabulary can be imposed on the group, from which all words used in communication must be chosen. The kinds of words permitted should influence the kinds of concepts that develop, and so guide the behavior of a given group in new test situations. Thus, certain lacks in the vocabulary, as of words designating direction or words designating position, might restrict the range of ideas explored in the learning process or lead to the invention of compensating idioms, while the presence of certain key words might predispose to certain hypotheses. If, for example, the word "sum" occurs in the vocabulary, there may be a strong tendency to see the grids as number puzzles, to which certain sums are the key. It so happens that five of the seven vertical sums in the examples given are multiples of three and the subject may well be led to use this sum in forming patterns, although no sum is a true key to the task. On the other hand, since certain sums might be statistically favored in the correct patterns, preoccupation with irrelevant criteria may nevertheless result in improved scores. Further, groups may be allowed to ask for additional words now and then, or even to "buy" them for score points, etc., to see whether ideational needs become acute. Or, groups might be permitted to develop their own vocabularies, subject only to size limitations. Again differential developments and their effects on the hypotheses formed, and thereby on the degrees and rates of adaptation achieved, would be at the center of interest. Finally, artificial languages only might be allowed.

There have been some preliminary ex-

plorations with the invention of artificial languages. A short passage (of some twenty words) is read to the subject, who is instructed to record the message, in any symbols he may choose to invent (any conventional or well-known code is disallowed), so that he can reproduce the message at will. Among several dozen subjects, all but one, who attempted a phonetic system, devised ideograms. Ideograms were often combined, as is done in Chinese, and conventional notations for successive abstractions or variations of an ideogram were invented. Thus, two ovals (the hemispheres) having been used for "brain" at one time, the word "idea" at once became an arrow into the ovals.[19]

If now the content of the messages is so biased as to favor certain classes of ideograms and their combinations, differential developments of "language" can be achieved through the subjects' own habits. These biases probably will be more stable than those imposed through a ready-made vocabulary. Thus equipped with a language, seemingly of his own making, the subject is instructed to teach this language to his group. It is expected that the internalization of the particular language will become more complete in the process of teaching it.

Suppose now, by doing all we can do to encourage "cultural differentiation" among several such groups, we "breed the cultures," i.e., establish lines of descent for the several groups through the process of replacing individuals. We may expect to see a process schematized as follows:

$$F_0(L) \rightarrow F_1(L) \rightarrow F_2(L) \rightarrow \cdots \quad \text{or}$$
$$F_0(L_0) \rightarrow F_1(L_1) \rightarrow F_2(L_2) \rightarrow \cdots$$
$$F'_0(L') \rightarrow F'_1(L') \rightarrow F'_2(L') \rightarrow \quad \text{or}$$
$$F'_0(L'_0) \rightarrow F_1'(L'_1) \rightarrow F'_2(L'_2) \rightarrow \cdots$$

where the F's indicate the "generations" and the L's their associated languages. The total language resources could be kept fixed over the generations (L or L') or be allowed to shift along with other elements. (L_0 to L_n or L'_0 to L'_n).

Degrees of adaptation are indicated by the scores achieved. Adaptations may be to the "actual" environment (the task) or to the "secreted" cultural environment (sets of established habits). There may be cultures characterized by bold exploration (trying out new ideas), passing through troughs on the adaptation curves to reach higher peaks but sometimes becoming "extinct," and there may be others clinging to safe conservatism, unchanging and mediocre. There may be lines of descent characterized by easy adaptation to drastically changed situations (new shape or size of grid) and others quite rigid. We may observe the appearance of innovators, who may become "martyrs" or "leaders" toward success or failure. We may observe invention and possibly rationalizations against the acceptance of the invention, especially in the cases of strong traditional bias. We may observe role differentiation, (e.g., the assignment of the job of "educating the young"), and even the "education" process itself.

Finally, when several "strains" based on fundamentally different "philosophies" have been bred to some degree of constancy, they can be "crossed." An individual, for example, may be transferred from one group to another and either allowed to bring his language with him or not. If he does bring his language, he may teach it to the others or forget it and learn the new one. But, having learned the new language, he may still retain the orientation imposed by his old one. The variations one can introduce are legion; so at this stage we offer an experimental program, not a description of definite experi-

[19] Incidentally, certain ideograms are chosen so commonly for particular words—as a group of waving vertical lines for fire and a similar group of waving horizontal ones for water—that deep cultural or "racial" elements may well be involved. Experiments with different cultural or age groups will be especially interesting in this connection. At least, individual characteristics have been picked up in terms of ideograms chosen. Thus, the word "Greece", was represented by such symbols as: a short sword, a few columns supporting an architrave, a Grecian nose, and the like. Two individuals, who used an outline map of Greece, turned out to have very powerful, almost eidetic, visual memory. The technique of having subjects create drawings to fit concepts, rather than fit stories to presented figures, might be developed into an inverse type of projection test.

ments. The program can be carried out only stepwise, utilizing the successes, failures, insights, and disappointments inherent in any experimentation to guide subsequent steps. Nevertheless, the following considerations concerning possible ranges of these variables and their importance may be useful guides.

The *kinds* of subjects chosen for participation might be highly important. One could start with college students, as a reasonably characteristic sample of adults enculturated in our culture, but would hope eventually to include adults of a different culture, as the Navaho Indians; urban children, at different stages in the course of enculturation in our own culture; and schizophrenic patients, presumably with the standard enculturation but with abnormal individual reactivities to it.

The *number* constituting a given group must be large enough so that the rotating of individuals does not completely change the entity. This would demand certainly three, and probably four, participants. Conversely, since experiments indicate that group pressure toward conformity by an individual is already fairly maximum at a ratio of three to one, there would seem to be little point in increasing the numbers and complexity of the group much beyond six. Pending further empirical evidence, therefore, the groups will start with four or five individuals.

The *length of participation* of a given person and the rate at which individuals rotate in and out of the group are also important variables for experiment. An important factor to control is contamination of incoming individuals by premature exposure to gossip on the part of those who have been through the group. This could be completely eliminated if it were possible to run a continuous experiment within twelve, or even twenty-four, hours, insuring that those individuals rotating out of the group do not have access to the persons already selected to rotate into it. Other devices would be: to select subjects at a military staging area so that men rotate out of the group just as they are sent overseas; to use bedridden (orthopedic) patients in a large hospital, with communication entirely by messengers concerned with the experiment; even to use members of different branches of the same company located in different cities. Aside from contamination, considerations concerning rotation would involve: the degree of "learning" desired before transmission, the established conditions of transmission, the desired pressures toward conformity, and the like.

A related variable is the *conditions of interaction between two independently developed groups*. The groups might be mixed *in toto* or reshuffled into two new groups, a single representative of one might be introduced into the other, the ejects of both groups might become the entrants to a third mixed group, communication might be complete by personal interaction or limited to prescribed communication channels and terms, and so forth. Although here, also, some prior analysis is possible, probably much experimentation will be needed to determine the actual effects of these variables.

Along the same lines, the *patterns of interaction and action* within the group at any one time present several parameters for decision—and for empirical guidance. Should individuals be allowed to interact freely and, in effect, make a series of joint decisions, as for moves in a game; should each member of the group make a "move" simultaneously, or in a rotational order, or by self-selection (as raising a hand in a quiz performance to volunteer the answer); and, in each of these cases, should "moves" be at fixed identical time intervals, or with limits as to total time (as in a chess tournament), or with only successive temporal units noted. Moreover, aside from the actual making of "moves," similar questions arise as to permissible communication between members of the group. Should this be allowed after each move or only after each task has been completed and different ones are compared, freely within the group or along specified channels, etc. As an especially important subhead here, already considered sufficiently, is the question of the words or written symbols that should be permitted in *communication*.

Again related, is the problem of *trans-*

mission of the group culture to the new arrival. This might include material transmission, the group being allowed actually to produce or modify objects, or the scoring procedure, as indicated above; or semantic transmission, by verbal or written communication or by less explicit participation and observation. An especially important variable here would be the extent to which the ideas transmitted are explicitly formulated into words and rules, and whether or not the transmitting individual has been preinstructed as to the need of formal transmission of what he has learned.

The extent of the *pressure upon the newcomer* to conform to the group culture, as contrasted with his freedom to introduce innovations, can be influenced in several ways. If the "environment" is very severe—so that the group "dies" if its performance falls below a predetermined level—the premium will be on conservation rather than experimentation. If the group has been forced to create a rather extensive esoteric communication system, the newcomer can hardly exert an influence until he has become thoroughly acculturated, by which time he is less likely to introduce a completely fresh viewpoint. The rules chosen (by the experimenter) concerning interaction within the group could also favor freedom or conformity on the part of the newcomer. Finally, if the actual performance of the group has been allowed, by the scoring device chosen (see above and later), to influence the excellence of the performance, a deviation from the established ways of operation is much less likely to constitute an improvement than if such weighting has not been introduced.

As already indicated, the *scoring rules* can markedly influence the direction of change of the group. Two obvious factors are: whether individual performances are scored and totaled in some way, or whether a single group performance is scored; and whether both positive and negative factors are used—as introducing a penalty for sufficiently poor performance, or not. More subtle, and potentially highly important, is whether scoring is done on a fixed, continued, and uniform scheme—in which case the group must evolve towards a set goal; or whether the actual scoring principles are allowed to shift over the passage of time, by changing the weights given particular components in terms of what the group has actually done. If such a slanting of the score by group performance is introduced, the goal is not fixed but is to a desired degree generated by the group itself. As mentioned, such a situation would place greater restraints on an incoming member of the group. Such a procedure, moreover, would introduce complications when two groups were subsequently mixed, since each would be scored on a different basis. Possible difficulties here can be overcome in two ways. One is simply to grade each action taken by a given member of the mixed group in terms of the scoring level that pertains to the group from which he came. This would work quite smoothly if members of previously different groups were allowed to act alternately. The second method would be gradually to bring the scoring weight back to the initial conditions, so that both groups would end, as they started, with the same scoring rules. The initial period of group slanting would presumably be effective in fixing one or another tradition or hypothesis in the group during its cultural evolution.

Closely related to the scoring rules is the information given the group regarding its performance, the *scoring feedback.* This could range from essentially none, which would correspond to minimal selection pressures; through a statement merely of better or worse, perhaps most simply handled by allowing each run to consist of two performances and the superior one indicated; to the giving of a precise numerical score or even exhibiting the correct solution. A further variable would be the timing of the feedback, whether after each run, series, or entire experiment.

Another closely related variable is the *nature of the reward.* This might be in terms of individual performance and, indeed, could influence the continued participation of the individual in the group, if rotation out were in terms of performance rather than of time. Or, of course, partners or

other subgroups could be rewarded, or the reward limited to the whole group on its total performance. Clearly, the degree of cooperation and cohesiveness would be strongly influenced by the reward pattern chosen. Moreover, this same factor can be extended, as between a given generation and following ones, by rewarding the individual or group in terms of immediate performance of the participants or in terms of subsequent performance of those taught.

Finally, but most important of all, is the *nature of the task* assigned to the group. The nature of the task will generally determine the *initial "set"* toward it—tasks resembling puzzles tend to act as invitations to match wits with the experimenter. But the same problem may be presented quite differently, as in terms of an esthetic evaluation of a design where the subject is not expected to reason out an answer, and his reactions might also be quite different.

In any event, it is possible to emphasize the generation of hypotheses or the generation of communication devices or both. The primary aim is to have a group generate a hypothesis or a group of related hypotheses concerning a task in such a way that it is possible experimentally to influence the direction in which these hypotheses will proceed. Perhaps the most restrictive kind of hypothesis, because the most logically determined, would be that presented by a number series, as discussed. Less restrictive, are such tasks as completing a picture or extending a two dimensional graph or selecting a set of cards. Still less related to a logical solution are tasks involving esthetic evaluation. Finally, at the other extreme, there may be no task as such; the group may be asked to do nothing but spend some time in a room talking. The "key" to the score would be an arbitrary list of actions, assigned positive or negative values, e.g., crossing the legs or other gestures, walking, standing, amount of talking, occurrence of certain words in the conversation. If the group is constantly informed of its current score, it will learn to improve it in the course of accidental variations of its behavior, without, however, being aware of just what is being learned. Here learning is devoid of all intellectual content. Yet if the group is required to pass on to successors what it has learned, it must rationalize, that is assign intellectual content to, its behavior. It would be interesting to see what course "cultural evolution" takes under these circumstances.

It would be surprising if the actual development of the program does not differ radically from that proposed at this preliminary stage; indeed, if the program is undertaken by more than one group of workers, they will almost certainly go in different directions. For a team of investigators also forms a "society" and develops a "culture," with its characteristic biases, adaptations, techniques, and "superstitions," and the business of breeding cultures and crossing them can be applied to the investigators as well as the subjects.

To sum up, the program offers methods of laboratory control of the following variables, thought to be important in the evolution of human cultures:

1. The world and its problems—the environment. By varying the scoring procedure the environment can be made mild or harsh, constant or erratic.

2. A society of individuals, self-sufficient with respect to the environment. The size of the society, its birth and death rates, and its composition with respect to the background of the members can be controlled.

3. The recording of experience in a language, subject to experimental control and observation.

4. The communication process within the society.

5. The transmission of knowledge and of environmental modifications to future members of the society.

Some questions which arise at this stage of conceptualization of the program are:

1. Can "strains" of these miniature cultures be "bred"; can a given "micro-culture" be kept stable in spite of the constant replacement of the members of the society? What are the factors making for stability or instability of the micro-culture? The question pertains not merely to the development of different habits in different

individuals under different conditions, but rather to the possibility of determining "cultures" which are a further abstraction removed from the "societies" in which they arise.

2. If different distinct strains are obtained, can their traditions be correlated with the languages in use, with the complexity and capriciousness of the environment, with other controlled variables?

3. Will rate and direction of development of the micro-cultures, as reflected in their performance of languages, show regularities depending on the experimental conditions? Is senescence possible?

4. Can one distinguish the "rational" and the nonrational factors in the repertoire of ideas which a group acquires about its task? Can the usual aberrations be observed ("cultural neurosis", xenophobia, what not)?

5. What happens when micro-cultures are crossed? Can resulting changes be correlated with the observed changes in the "technology"? Are changes occasioned by "immigrants" different from those arising within the "indigenous" groups?

6. Are there counterparts of the fate of biological species, such as fortuitous variations which under certain conditions lead to new and higher peaks of adaptation and under other conditions to extinction?

CONCLUSION

As stated at the beginning of this article, we are offering a plan for research more than any research results or even any particular experimental design. A bare minimum of actual testing has been carried out; but, although we plan to engage in an extended experimental program along the lines indicated, it seems wise to present this gleam in our experimental eyes without waiting until definitive results of the conception are available. The approach seems sufficiently novel so that major improvements could easily result from a wider consideration of the proposed experiments; and we warmly invite critical discussion and further positive suggestions, which may prevent unnecessary or useless experiments or suggest far better ones than have as yet occurred to us. Moreover, if the proposed techniques do prove feasible, the opportunities for fruitful research are almost unlimited. Similarly, the viewpoint here developed concerning the evolution of language might influence the observations and analyses of professional linguists. We would hope that many others would find investigations along one or another line indicated to be of sufficient promise and interest to pursue some ramification of them. Such fellow investigators we are eager to encourage to participate from the start.

REFERENCES

1. Boas, F. The classification of American languages. *Amer. Anthropologist*, 1920, 22, 367–376.
2. Bonfante, G., & Sebeok, T. A. Linguistics and the age and area hypothesis. *Amer. Anthropologist*, 1944, 46, 382-386.
3. Casagrande, J. B. Comanche linguistic acculturation II. *Int. J. Amer. Linguistics*, 1954, 20, 217-237.
4. Chao, Y. R. Review of G. K. Zipf's *Human Behavior and the Principle of Least Effort, Language*, 1950, 26, 394–401.
5. Gerard, R. W. Higher levels of integration. *Science*, 1942, 95, 309–313.
6. Gerard, R. W. Levels of organization. *Main Currents*, in press.
7. Glass, B. Pseudoalleles. *Science*, 1955, 122, 233.
8. Greenberg, J. H. Language change. In C. Osgood (Ed.), *Psycholinguistics*. Memoir 10 of the *Int. J. Amer. Linguistics*, 1954.
9. Hall, E. T. and Trager, G. *The Analysis of Culture*, Washington, 1953.
10. Harris, Z. S. Review of *Selected Writings of Edward Sapir in Language, Culture and Personality, Language*, 1951, 27, 288–333.
11. Hoijer, H. Linguistic and cultural change. *Language*, 1948, 24, 335–345.
12. Jacobs, M. Historic perspectives in Indian languages of Oregon and Washington. *Pacific Northwest Quart.*, 1937, 28, 55–74.
13. Jakobson, R., & Lotz, J. Notes on the French phonemic pattern. *Word*, 1949, 5, 151–158.
14. Kerr, W. E., & Wright, S. Experimental studies of the distribution of gene frequencies in very small populations of *Drosophila Melanogaster*. *Evolution*, 1954, 8, 172–177.
15. Kluckhohn, C. The study of culture. In D. Lerner & H. Lasswell, *The policy sciences*. Stanford: Stanford University Press, 1951. Pp. 86–102.
16. Landau, H. G., & Rapoport, A. Contribution to the mathematical theory of contagion and spread of information. *Bull. math. Biophysics*, 1953, 15, 173–183.
17. Mandelbrot, B. An informational theory of

the statistical structure of language. In W. Jackson (Ed.), *Communication theory*. New York: Academic Press, 1953.
18. Martinet, A. Function, structure, and sound change. *Word*, 1952, 8, 1–32.
19. Rapoport, A. Response time and threshold of a random net. *Bull. math. Biophysics*, 1952, 14, 351–363.
20. Rapoport, A. Spread of information through a population with sociostructural bias: I, II. *Bull. math. Biophysics*, 1953, 15, 523–533; 535–546.
21. Reichard, G. A. Significance of aspiration in Navaho. *Int. J. Amer. Linguistics*, 1948, 14, 15–19.
22. Stetson, R. H. Bases of phonology. Oberlin: Oberlin College, 1945.
23. Swadesh, M. Review of M. Cohen's *Le Langage*, *Int. J. Amer. Linguistics*, 1954, 20, 342–345.
24. Weinreich, U. Languages in contact. New York: Linguistics Circle of New York, 1953.
25. Whyte, W. The transients. *Fortune*, 1953, 48, (July) 84–89.
26. Wright, S. Evolution in Mendelian populations. *Genetics*, 1931, 16, 97–159.
27. Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proc. Sixth Int. Congress Genetics*, 1932, 1, 356, 366.
28. Wright, S. On the roles of directed and random changes in gene frequency in the genetics of populations. *Evolution*, 1948, 2, 279–294.
29. Zipf, G. K. Cultural-chronological strata in speech. *J. Abnorm. soc. Psychol.*, 1946, 41, 351–355.
30. Zipf, G. K. Human behavior and the principle of least effort. Cambridge, Mass.: Addison-Wesley Press, 1949.

(Manuscript received October 15, 1955.)

Samuel Rogers said:

Wordsworth and myself called on Coleridge when he was living at Gillman's. We sat with him two hours, he talking the whole time without intermission. When we left the house, we walked for some time without speaking.

"What a wonderful man he is!" exclaimed Wordsworth.

"Wonderful, indeed," said I.

"What depth of thought, what richness of expression!" continued Wordsworth.

"There's nothing like him that I ever heard," rejoined I.

Another pause.

"Pray," inquired Wordsworth, "did you precisely understand what he said about the Kantian philosophy?"

"Not precisely."

"Or about the plurality of worlds?"

"I can't say I did. In fact, if the truth must out, I did not understand a syllable from one end of his monologue to the other."

"No more," said Wordsworth, "did I."

From Irving J. Lee, **Customs and Crises in Communication,** Harper and Brothers.

THE EFFECT OF EXPERIENCE ON A DETERMINATE DYNAMIC SYSTEM

by W. Ross Ashby

Barnwood House, Gloucester, England; and the Center for Advanced Study in the Behavioral Sciences

The kappa in Phi Beta Kappa stands for "kubernetes," a Greek noun related to the English adjective "gubernatorial." It refers to governing or executive functions. (ΦBK means "philosophy, the governor of life.") The name for the new science of governing mechanisms, cybernetics, is derived from the Greek "kubernetes." This year the first text in this field will appear, written by the author of this article, who also wrote "Design for a Brain." Here he gives a preview of some of the ideas he develops in his new book, particularly how it happens that the state of certain systems can with the passage of time be more dependent on what their surroundings have been like than on their initial states. He relates his theoretical concepts to observations about how personality and behavior patterns develop as human beings mature.

In the last decade, science has undergone a major extension. The earlier science, led by physics and chemistry, concerned itself chiefly with such systems as were either simple or could be made simple by suitable transformation. On the whole, the study of systems that were intrinsically complex was avoided; the chemist preferred to study myoglobin rather than steak, and amino acids rather than proteins. In psychology and the social sciences the aim tended to be similar: analysis to the simple so as to avoid the complex.

The new movement in science might perhaps be dated from the '20's, when R. A. Fisher, studying the relation between the treatment of agricultural soils in one year and the responses of the crops in the next, realized consciously that he was dealing with a system whose complexity could no longer be set aside as a minor matter. The century-old strategy—alter only one factor or variable at a time—was simply not possible; for an alteration of, say, the soil's content of lime in the autumn would lead to alterations in innumerable other variables long before the end of the experiment was reached. Fisher realized that a fundamentally new strategy was called for, and he had the skill to develop one. Today the strategy of varying the single factor still holds an important place in scientific method, but it is used only when the system is fairly simple: when the system becomes complex we have a strategy specially suited to the complexity.

The methods developed by Fisher set many biologists thinking. Here was a method that accepted the soil as a living, reactive, and highly complex dynamic system, and yet succeeded in winning from it quite simple answers to simple questions. Could not similar methods be developed for use in the other sciences?

Some might regard Fisher's work simply as something to be added to the body of statistical techniques already available; but this view seems superficial. Fisher found that he was forced to go back to fundamentals, and in particular he was forced to clarify and develop the general theory of "information." He found that as this concept became clearer, so were the new methods deduced as *necessary* consequences of the basic concepts.

Ten years later, Shannon and Wiener (6) were considering various problems of telephonic coding, in which the complexity was

intrinsic. They also found that only when the concept of "quantity of information" had been refined and made objective could they develop an adequate theory of coding and communication.

We who follow can now begin to see what is what. We are beginning to see that complexity in major degree is not an insuperable barrier to knowledge, provided it is tackled in the appropriate way. The way *not* to tackle such a system is by analysis, for this process gives us only a vast number of separate parts or items of information, the results of whose interaction no one can predict. If we take such a system to pieces, we find that we cannot reassemble it! What *is* the best strategy is not yet clear, but many of us are convinced that suitable and practical strategies exist, and that they are now in process of being discovered. It seems almost certain that in these strategies a central part will be played by the concept of information.

The problem of the Black Box

Among the problems to be tackled by those who would study the complex system is that of the Black Box. It arose in electrical engineering, but its range of applicability is far wider. In front of the investigator is a box with electrical terminals for input, to which he may apply such shocks, potentials, and other stimuli as he pleases; and a set of terminals for output, at which he may observe what he can. The box, for whatever reason, cannot be opened. Any knowledge that he can obtain of its contents is to be obtained only by deduction.

The problem, in principle, occurs ubiquitously. It occurs when a mechanic, facing a troublesome engine, makes a few tests to see if he can deduce the cause without opening it. It occurs when the biochemist, facing the complex enzymatic systems of the living cell, applies mild poisons, observes the results, and makes deductions about the molecules that he cannot observe in actual action. And it occurs when the psychologist, trying to understand and predict the behavior of a living rat, puts it into a maze with stimuli, observes its responses, and then tries to make deductions about what is going on inside its skull.

Faced with the problem of the Black Box, the student of behavioral science in general should beware of becoming immersed in the study of a particular box. The number and types of Boxes are so vast that behavioral science can remain a unity of manageable size only if its subject is confined to those properties that are generally true. Thus a question, relating to the Black Box, that behavioral science might well tackle is: what internal properties of the Black Box are deducible and what are *fundamentally not deducible*? Clarification on this point might enable workers in the various branches of the behavioral sciences to select their aims with a better relation to what is achievable.

The logic of mechanism

Before such a question can be tackled, we need a better logic for arguing about mechanism in general. At the moment, most arguments about mechanism are based on experience with actual machines; and as the world's machines (in the general sense) do not form a noticeably orderly set, so are the arguments apt to be scrappy and lacking in cohesion. Can a uniform Theory of Machines be built?

In the author's opinion it can. It is partly the purpose of this paper to introduce the subject and to give one example of it in operation. A more extensive account has been prepared and is being published elsewhere (2). Its basic concepts are drawn from the theory of sets, as developed by the French school who publish under a pseudonym (4). These concepts have already been shown, by that school, to be fundamental in all branches of mathematical—i.e., of exact—thought, so we need have little hesitation in building on the same foundations. Fortunately, comprehension of these concepts calls for no advanced mathematical technique, only the sort of close reasoning that is required for discussing a position at chess.

Such a Theory of Machines should no more be restricted to the properties of actual machines than is the Theory of Surfaces restricted to those surfaces that are constructi-

ble in three-dimensional space. A logic of mechanism is required only to be consistent within itself, and to provide a conceptual richness sufficiently comprehensive for all actual machines. Its special purpose is to enable arguments about machines to be conducted rigorously, so that any particular argument, i.e., chain of deduction, can be shown unambiguously to be either valid or invalid.

Its basic concepts will now be described (1, 5).

The machine with input

It is assumed that the system, at each moment of time, is in one and only one of a definite set of states S_1, S_2, S_3, $\cdots$. (The set may be finite or infinite; the states may be finitely separated or infinitely close; and the states may be specified by vectors with any number of components. Thus the range of special cases is much wider than might at first appear.)

A real system is called *absolute* (over some given range of observations) if the observations show that each state, whenever it occurs, is always followed by one state invariably. Thus if, on one occasion, S_5 has been observed to be succeeded by S_2, then the system is absolute only if, on every other occasion when S_5 occurred, it was also always followed by S_2. Similarly for the other states.

Any system that behaves in this way can have its behavior specified compactly by a transformation which simply gives, for each state, its successor. The transformation may be written down in a variety of forms, which differ only in their convenience for particular systems.

The basic form simply specifies directly the set of states and the successor of each. Thus, if the states possible are a, b, c, d, and e, the transformation might be:

$$T: \downarrow \begin{matrix} a & b & c & d & e \\ e & b & e & b & c \end{matrix}$$

The arrow means that a is always succeeded by e, b by b, and so on.

If the states are vectorial, i.e., compounded from components, the transformation will usually show how each component changes. Thus if each state were defined by the numerical values of three components x, y, and z; and if the succeeding values are x', y', and z', so that the state (x, y, z) is succeeded by the state (x', y', z'); then the transformation will be of the form

$$\left.\begin{aligned} x' &= f_1(x, y, z) \\ y' &= f_2(x, y, z) \\ z' &= f_3(x, y, z) \end{aligned}\right\}$$

For the system to be absolute, the sole restriction on the functions f_i is that they be single valued.

If x' is only infinitesimally different from x, i.e., if x is a continuous function of time, then $x' = x + dx$, and the transformation may be written:

$$\left.\begin{aligned} \frac{dx}{dt} &= g_1(x, y, z) \\ \frac{dy}{dt} &= g_2(x, y, z) \\ \frac{dz}{dt} &= g_3(x, y, z) \end{aligned}\right\}$$

Again the sole restriction on the functions g_i is that they be single valued. Bertalanffy (3) has given a discussion of these forms, and has shown how they represent all of the commonly accepted properties of systems. The differential form is that most commonly used in mathematical physics, but the basic form is the most general, i.e., the least restrictive on the system.

If an initial state be given, repeated application of the transformation generates a *trajectory*. (This repeated application corresponds, in the real machine, to the march of time, step by step. Thus, if the transformation first mentioned is T, and the initial state is a, then $T(a) = e$, $T(e) = c$, $T(c) = e$, and so on. $T(T(a))$ can be written $T^2(a)$. Then the sequence a, $T(a)$, $T^2(a)$, $T^3(a)$, $\cdots$ gives the trajectory $a, e, c, e, c, e \cdots$.

A *state of equilibrium* (resting state) under a transformation U is one that satisfies $U(x) = x$, so that, as time goes on, the system, if at this state, will stay there. T above has only b as state of equilibrium.

The concept of the absolute system gives rigor to the intuitive concept of the system that is determinate and *isolated*. Such, for instance, is a good clock; for if we know the length of its pendulum, the state of its spring, the positions of its wheels, etc.—i.e., know its "state"—then the trajectory that it will follow is determinate—provided no outside factor (a draft, say) comes to affect it.

A more general concept is that of the *machine with input*. This system has similarly a set of possible internal states, but it also has an input that, at each moment, can be in one of a definite set of input states $\alpha, \beta, \cdots$. Further, the transformation that applies at each moment is a function of the input state at the moment. For example, a machine with input might have internal states a, b, c, d, and e. Its input might be able to take states α, β and γ. A possible set of transformations would then be:

		Internal State				
	↓	a	b	c	d	e
Input state	α	a	a	d	d	a
	β	e	b	e	b	c
	γ	a	b	c	b	e

Then, if its input is at α, it behaves according to the transformation

$$\alpha: \quad \begin{array}{c|ccccc} \downarrow & a & b & c & d & e \\ & a & a & d & d & a, \end{array}$$

and similarly for the other input states. This formulation says that if, at a given moment, the system's internal state is c, and its input state is α, then the system will change from c to d. Had the input state been β it would have changed to e. The fact that each row and column intersect in a single element corresponds to the fact that the system's behavior, given its internal and input states, is determinate.

By repetition, the trajectory can be traced for any sequence of input states, which are assumed to change in step with the system. Thus the initial state c and the input sequence $\beta, \gamma, \alpha, \alpha, \beta$ gives the sequence of states (the trajectory) c, e, e, a, a, e; similarly the input sequence $\beta, \gamma, \beta, \alpha, \beta$ (differing from the former at only one point) gives the trajectory c, e, e, c, d, b. There is, thus, an unambiguous representation of a system receiving an input and giving an output, and imposing something of its own nature on the relation between them. It also gives exact expression to the century-old idea that a "machine" is that which, if in a given state internally and in given conditions externally, will do one thing and one only.

It should be noticed that if the input is fixed at some value, the "machine with input" becomes an "absolute system." Thus, in the example just given, if the input is fixed at β, the machine becomes identical, so far as behavior is concerned, with the absolute system T described earlier.

Variety in Machines

We must now introduce the concept of "variety"; it is implicit in the concept of information, but is more primitive. Given a set of elements, the variety of the set is the number of *distinct* elements in it. Thus the set of eleven letters:

$$d \quad c \quad c \quad a \quad c \quad a \quad a \quad d \quad a \quad d \quad c,$$

if regarded as a collection of individual letters, has a variety of three. In this paper we shall be concerned exclusively with the variety that may occur within some set of a machine's states.

How is the concept of "variety in state" to be applied to a given machine, when any real machine can only be in one state at a time? Variety may occur in several ways.

First, there may in fact be many replicas of the machine, ("mass-produced"), identical in structure (i.e., in their transformations) but at different states; as ten identical watches might show at a given moment various times by their hands.

Secondly, one machine may be considered on various occasions; if the distinctions between the occasions are irrelevant to the main interest, the one machine at various occasions may be regarded as equivalent to various replicas acting simultaneously.

Thirdly, some observer may wish to consider, not the state the machine *is* at but the set of states it *might be* at, the size of the set of possibilities corresponding to some incompleteness in his knowledge about the sys-

tem's present state. (Full knowledge restricts the set to one state; zero knowledge enlarges the set to include all possible states.) This is the situation in the statistical mechanics of a gas; the observer cannot discover the gas's state with microscopic exactness, so he is forced to consider a whole set of possibilities.

In many contexts the three situations are equivalent. In this paper I shall use a terminology appropriate to the first, as I find it conceptually to be the easiest. (Our conclusions will be found to be equally valid for all three).

Suppose then that we have before us an absolute system, specified by a transformation T. Suppose also that we have replicas of it, that behave similarly, by the same transformation. Suppose that some are at state a, some at b, $\cdots$, and some at e. When they all change over one interval of time, all those that are at one state x must go over to one state y; no scatter can occur, and the variety cannot increase. On the other hand, those at w and those at x may both go to y, so *variety can be lost.*

At the next step the argument holds just as before. Thus, *in a set of identical absolute systems, the variety, as time goes on, can only fall, or perhaps stay unchanged.*

The same proposition holds true for a set of identical machines with input, if all the inputs are throughout held constant at the same state; for then the case is identical with that just described.

Next, we can notice that if the variety in state, in a set of identical absolute systems, falls to a minimum (as it must eventually, for it cannot fall below unity) this minimum need not be unity itself. Thus if T is applied repeatedly to a set that starts with a variety of five as in the first example above, the variety falls to three (states b, c, and e), and then remains constant. (The same applies to a set of machines with input (as exemplified) with all inputs fixed at β.)

Suppose now that, in this set with variety sunk to three, all inputs be changed to α. As is easily verified, b, c, and e go respectively to a and d; so the variety sinks to a new, and lower minimum. Thus in a set of (replicated) machines with inputs, in which the variety (under constant input) has already sunk to a minimum, *change of all the inputs to a new value may allow the variety to fall further.*

What this would mean can be shown by an example. Suppose we are considering the same machine with input that we have discussed before, suppose we face a particular one, and suppose we do not know which state it is at. The possible states have variety 5; this number also measures our uncertainty. We now put in some input sequence such as $\beta, \gamma, \alpha, \gamma, \alpha, \beta, \beta, \alpha$. At the end of it we find the machine to be in state d. Now the reader may easily verify, as he knows the machine's set of transformations, that if this input sequence is given to a set of such machines, some of which start at a, some at b, etc., so that the initial variety is 5, then as step succeeds step the variety over the set goes through the values 5, 3, 3, 2, 2, 1, 1, 1, 1, (always either holding its value or decreasing). At 1 it must remain; and the fact that the variety is 1 means that whatever the initial states of the 5 possible trajectories (for the particular input sequence) their states after the sequence must all be the same, (they are in fact all at d). So if we do not know what the initial state is, but we do know that input sequence $\beta, \gamma, \alpha, \gamma, \alpha, \beta, \beta, \alpha$ has been applied, then we can predict that the system is now at d, *regardless of what the initial state was.*

Suppose now that another input sequence is used, say $\gamma, \beta, \gamma, \alpha, \beta, \gamma, \gamma, \alpha$. Test shows that on this sequence too the variety happens to fall to 1, and that, whatever the initial state, the system is finally at state a. The distinction between d and a (as final states) is clearly attributable wholly to that between the two input sequences, and not at all to any distinctions that may have existed between the initial states. Thus the following rule emerges:

If the input sequence is long, the state of the machine at the end of the sequence depends more on which sequence was put in and less on which state the machine was at initially. Further, the longer the input sequence, other things being equal, the more does the final state depend on the input rather than on the initial state.

This statement is not in properly rigorous form, but that is not important at the moment, for the rigor could easily be supplied. What is important to us as students of behavioral science in general is the general nature of the phenomenon.

The phenomenon, once demonstrated, can be looked at from various points of view:

1. Suppose the machine is started from some state that is unknown to the observer, and that at the t^{th} step he observes what state it is at. If t is small (especially if t is zero!) he gains some information about which state was present initially. But if t is large, he gains little or no information about the initial state, but he does gain information relative to which sequence has been applied at the input.

2. The same phenomenon can be described in a more picturesque form if we remember that the machine's set of states, being limited to 5, allows it to contain information, in respect of its state, only to the amount of $\log_2 5$, i.e. 2.3 bits. At first, this quantity of information is wholly relevant to its initial state; later, however, the state it is at gives more and more information about which input sequence it was subjected to, and less and less that is relevant to which state it was at initially. Thus we may say, picturesquely if hardly rigorously, that *information supplied to a determinate machine through its input eventually swamps information about the machine's initial state, and replaces it by information about which sequence was applied to the input.*

That our example should have shown such a swamping is not unreasonable. The set of states cannot contain more than 2.3 bits, while the variety put in may amount to much more. Thus, at each step the input may be (in our example) at any one of three states, giving a variety of $\log_2 3$, i.e. 1.6 bits; and eight such states, if independent, can provide eight times this, i.e. 12.7 bits. It is not surprising that the injection of 12.7 bits has tended, in our example, to swamp the 2.3 bits present initially.

To sum up with a statement that is more meaningful than rigorous: *every determinate system changes so that, as time goes on, its state is characterized more by the experiences that have come to it than by its state initially.*

Minor comments

Before we discuss the possible applications of this phenomenon, some comments may be advisable if minor misunderstandings are to be avoided.

It should be noticed that no assumption has been made restricting the system's states to those that vary continuously, either from state to state or as functions of time. The phenomenon occurs equally in those systems whose states change by discrete jumps in time and in those whose states vary continuously, and in mixed forms. It may therefore be properly applicable to biological systems, in which discontinuities are common.

The argument has been stated in terms of systems that are assumed to be discrete, both between state and state as they may occur in the machine, and as they are functions of the time. This form is of extreme generality, for the continuous forms may be obtained from it by passage to the limit as the differences or steps become progressively smaller. The discrete form has been used because, as shown so clearly by Shannon, it provides the simplest and clearest introduction. The attempt to begin with the continuous forms leads at once into considerable mathematical difficulties and subtleties, which are quite avoided in the discrete form.

In addition, no assumption has been made that states are related by some metric. It may therefore be applied both to those states that are specified by number and to those that are specified by verbal description (as the meteorologist specifies several types of cloud, or as the entomologist specifies several stages in an insect's life-history.)

In particular, the extremely restrictive hypothesis of linearity, so often introduced purely for its mathematical convenience, is wholly avoided. The formulation is thus natural to biological material, from this point of view.

It should be appreciated that the formulation above does not exclude the possibility that the inputs or the internal states may be

multiple and vectorial. Thus, if there are five channels for input, the "state" of the input will be a vector with five components. In the same way, the system's internal "state" may be specified by a vector with components. Such a "state" is shown vividly in the cockpit of an aircraft, where a number of dials give, collectively, a statement of the present "state" of the motors.

The theorem is clearly related to the concepts of communication theory as developed by Shannon. The "determinate system" of this paper is the same as his "finite state transducer," with the slight modification that its output is its internal state unchanged to new symbols. On the other hand, care is needed before the theorems of his method are applied to the conditions supposed to hold in our discussion. The concept of "variety" is much more primitive than Shannon's "entropy." The latter has meaning only when the sequence of events goes on forever and where the sequence is ergodic. It is, therefore, inapplicable to the cases we considered, in which the mechanism moved forward over a few steps, and perhaps moved to an equilibrium.

The theorems stated above gain considerably in applicability when one appreciates that the set of identical machines, considered above for clarity, need not be simultaneously existent. The same relations between input sequence and state will hold if we consider one machine at a set of occasions, which occur in succession during its lifetime. They will also hold if the machine is given, and then some set of states is specified by a given observer's lack of complete knowledge. When he discusses, or thinks about, the machine he will be forced to manipulate a set of states, for he cannot specify more finely.

The proposition does not attempt to dogmatize about what real systems must do; what it does is to show something of what is, or is not, a coherent argument about machines. It shows that if a real system has one property—that of being determinate and of having a single-valued transformation—then that same property, viewed from another angle, will show as an apparently different property—that of tending to lose variety under certain types of treatment at its input. It shows that the first property implies the second.

Applications

Perhaps its first application is to act as a guide to naive intuition, based on partial experience. Faced with a complex, but determinate, system, about whose present state there is some uncertainty, one may well feel that as some arbitrary sequence of actions is taken on the system, and as the input sequence gets longer and longer, so is the uncertainty about the system's state apt to become greater and greater. We have seen that this is not so; the naive expectation does not make sufficient allowance for the effect of the determinateness.

When we turn to the real world, phenomena that may be due to this process are innumerable; for many cases are known in which systems, initially unlike, become alike when given uniform treatment.

A possible example of the operation of the phenomenon is given when a number of boys, coming from varied homes and with varied histories, and therefore somewhat heterogeneous in present nature, go to some institution,—a naval training school, say—that subjects them uniformly to a long sequence of impressive events. After a time it is commonly observed that the boys' way of behaving can be related more closely to the school they went to than to their previous histories. (Whether such a tendency is really due to the process described in this paper can be decided with certainty only by more detailed examination.)

Something of the same nature may be seen generally in the process by which the infant's cerebral cortex changes from a system of neuronic connections that have been determined originally by trifling local details of blood supply, ionic conditions, etc., and therefore unrelated to events elsewhere. After the infant has grown up, however,—after it has received twenty years of education as input—its cerebral condition (and consequent behavior) is observed to depend more on which input sequence came to it

than on its original state. Again, it is not impossible that the process described may be in operation.

In the past, such observed phenomena have often been regarded as depending on some complex and very special mechanism. Often a variety of special mechanisms has been hypothesized for the various forms of the phenomenon, each invented *ad hoc*. What this paper shows is the possibility that no special mechanism need be invoked—that the presence of determinateness in major degree is sufficient as an explanation.

We have seen above that (roughly) determinateness implies ultimate loss of variety. We may use this proposition with the usual rule that if A implies B, then not-B implies not-A. Thus, with a given real system we might observe whether its variety falls; if it does not we may validly *deduce* that it is not determinate.

One of the outstanding questions of psychophysiology is to what extent the behavior of the living brain is determinate. The argument could be used as a test. For if it is found that a set of rats behaves, in defined conditions, with some scatter, but that after they have been put through some uniform treatment the scatter becomes greater, then this observation would reject the hypothesis that the system tested was determinate, in the sense defined here. The actual experiment would require many refinements, but the principle would be clear.

Throughout this paper no mention has been made of the particular parts that compose the system. It is sufficient for the argument if the system is determinate. We have thus seen an example showing how rigorous arguments can be conducted about systems in general, without answering the question "system of what?"

SUMMARY

From the assumption that a system is determinate many deductions, some of them by no means obvious, may be drawn.

One deduction, made in this paper, is that a determinate system, as time goes on, if affected by its surrounding conditions, will tend to change to a state that depends more on how the surrounding conditions have changed than on what state it was at initially. In the concepts of communication theory, information supplied at its input will tend to replace information present initially.

This proposition follows independently of particular assumptions about the working of the system's parts.

Two possible examples of its operation are given.

REFERENCES

1. Ashby, W. R. *Design for a brain*. New York: Wiley, 1954.
2. Ashby, W. R. *An introduction to cybernetics*, in press.
3. Bertalanffy, L. von. An outline of general systems theory. *Brit. J. Phil. Sci.*, 1950, 1, 134–165.
4. Bourbaki, N. *Théorie des ensembles*. (2nd Ed.) Fascicule de résultats. A.S.E.I. 1141. Paris: Hermann & Cie., 1951.
5. Riguet, J. Sur le rapport entre les concepts de machine de multipole et de structure algébrique. *C. R. Acad. Sci.*, 1953, 237, 425–427.
6. Shannon, C. E., & Weaver, W. *The mathematical theory of communication*. Urbana, Ill.: University of Illinois Press, 1949.

(Manuscript received October 10, 1955.)

Any two philosophers can tell each other all they know in two hours.

—Oliver Wendell Holmes, Jr.

COMMUNICATIONS AND SOCIAL CHANGE

by Richard L. Meier

Program of Education and Research in Planning, University of Chicago

Advances in communications theory in the last few years give the appearance of an important scientific breakthrough. They provide principles for translating into similar terms and quantitative units communication interactions both among human beings and among machines, as well as between men and machines. They also suggest aspects of a more inclusive theory of organization. Technology is being broadly influenced by these new developments—automation being a highly publicized and poorly comprehended example of such effects. This article outlines possible ways that the communications advances may alter such diverse aspects of social behavior as the flow of traffic, the use of mass media, and the development of personal values and esthetic standards.

IN THE PATH OF A BREAKTHROUGH

THERE could hardly be any development more interesting to observe than a breakthrough in science. It starts as a movement in an undefined shadow area, one of many barely discernible movements which occur in the advance of human knowledge. If one is fortunate, he may have before the event fixed his attention upon just that sector in which the breakthrough was made. Then he would have seen the same groping for fruitful concepts and the same meticulous refinement of categories and measurements as was to be observed elsewhere on the frontier. But then something very important occurs—it may be a totally new phenomenon or, instead, a level of generalization never hitherto achieved.

The individual or team responsible for the discovery realizes only in a vague, uncertain way what has happened, for there is always a lingering doubt, the possibility that an inconsistency may be uncovered at the last moment, so their imaginations are reined in. At a scientific meeting, when the potentials of the paper are evident to program chairmen, it is put into the key position in a series of highly specialized papers, sometimes even accorded a special position which permits a wider range of attendance. Occasionally the first public notice comes through a communication-to-the-editor in some of the more general professional journals.

The undercurrent of activity is much stronger. There exists a system of communication between the important research centers which uses letters, "discussion papers," drafts of articles, and personal visits as a means of transferring the very latest thinking. A new finding can spread through this network in a matter of days or weeks; a radical new conception, however, may take months because it is more difficult to convey informally. Thus it is not unexpected, considering the normal delays of publication, that the focal paper is confirmed, buttressed, and even extended within a matter of months after the first public announcement.

At this stage no more than a few hundred people are likely to recognize the revolutionary character of the new finding. Of them only a few score would be able to relate it to their own research activities. Nevertheless original contributions come cheap when it is possible to play with a new conception. It is not long before extensions and applications of the principle begin to show up in print. Bright young men, intent upon intellectual gratification and quick to see an opportunity also for gaining profes-

sional recognition, leap into the potential breakthrough with overwhelming ardor. Opportunists and curiosity seekers in other scientific disciplines are also attracted. Thus the trickle of papers rapidly becomes a flood. In this outpouring of creativity other stubborn bastions of paradox are overcome, new levels of inquiry are penetrated without apparent struggle, and so the breakthrough becomes reality.

Slowly, and with some reservations, social institutions begin to react. Research programs are modified, and new contracts are granted to explore the possibilities opened up. Then come the honorary lectures at leading universities, and the awards from the scientific and professional societies. Professorial chairs are proffered and new programs of study are instituted. The new conceptualizations are then well on the road to becoming an intellectual fad. Military officers are deputized to discover whether there exists some way of enhancing combat effectiveness through application of these new principles. Industrial firms begin to make up their separate minds about how far they should explore the utility of these ideas.

As the readership potential expands, the publishers will institute careful appraisals of the market. Is it good for a few popular articles, a book, a new series of books, or a gamble on a new journal? In their attempts to develop the market, many competitive efforts may reach publication at once. There will be a temporary surfeit of materials, and because of it the ordinary college graduate, whose work is wholly unconnected with these curiosities, becomes a little more aware of what lay behind the assertions, rumors, human interest stories, and sometimes even back page headlines, of a few years back.

What, then, *is* a breakthrough? I find no definition fully satisfactory, but an analogy sometimes helps. Imagine the filming of a hurricane, where the film has been reversed. . . . Out of the untidy debris and incomplete structures (representing empirical studies and small-scale theories), after tremendous turbulence and boiling of vapor, comes a view of calm stability, with orderly aggregations and movements. Organization is evident—activities have some observable increase in coordination at many levels of inquiry. The outcome is a new steady state, substantially more comprehensible than that which preceded it. This kind of breakthrough is a locally severe storm of *con*structive thinking in the world of ideas.

A breakthrough in science nowadays affects the decision making of a society long before it changes the technology, or the various means of making a livelihood. Whenever we set out to judge what are the social effects of a new development, we must first appraise the effect of the enlightenment upon the way people view their future and, in so doing, modify their behavior. This thesis may be more understandable when asserted as follows: As new opportunities and perils are introduced, wise and informed people will take account of them in their investments, plans, and other long-range decisions.

Communications theory breaks through

About eight years ago such a breakthrough was initiated in a small specialism now called information theory. Shannon (4) published in the *Bell System Journal* two papers demonstrating that the information sent over a noise-laden communications channel had the mathematical properties of entropy in thermodynamics and is related to the weight of evidence in the mathematical statistics of R. A. Fisher. Suddenly much of the puzzling work of previous years became meaningful. A rapidly expanding field that had about then come to be known as "cybernetics" was one of the first to adopt this new formulation. Shortly they became quite inseparable. Shannon's theory in a remarkably brief period became a central concept in the work upon automatic controls. However it was quickly apparent that this was no ordinary breakthrough, but would have a contagiousness seldom encountered in the history of science.[1]

[1] Probably no single person has been in a position to describe what actually happened after Shannon's publication. We know by now that a rather important intellectual chain reaction was

Following these ideas as they sift through the key educational and research centers can be quite illuminating. Among the scientists and engineers one can conveniently label most of the informal circles that exist for work and play, or just plain sociability, as being either highbrow, middlebrow, or lowbrow. (There are more subtle regional distinctions which can be made, and these may on occasion be important to understanding the diffusion of a set of new ideas, but in this instance they seem to be of minor importance.) Among the highbrows it is the quality of the idea that counts, its verve, imagination, and other rather shocking features that set the mind to racing in altogether new directions. The new theorizing about communications was very fascinating to these groups, because it permitted quite unexpected analogies to be drawn. Individuals in these groups attempted the most scandalous generalizations, after rejecting reasonable assumptions one at a time (a standard procedure among successful theorists), and have occasionally gained by this procedure some unusually revealing insights. The middlebrows are less impressed—they demand theories that are unambiguous and will guide experiments they have already planned. For them this new theorizing is just another convenient tool which becomes particularly useful when dealing with the new computers. Many of them consider it to be a nuisance for their purposes because it introduces another set of symbols and another jargon for principles which, as far as they were concerned, were already recognized and well understood. Their kinds of problems are rarely solved by a new "most general" formulation. Lowbrows seldom read the scientific journals likely to contain theoretical developments—they are narrow specialists whose greatest skill is often found at the tips of their fingers—so their circles are barely aware of the conceptual changes being formulated. Among them concepts associated with automatic controls may be felt to be a source of insecurity, if they are discussed at all, because the changes following in the wake of such ideas may render their specialty less important, and may even bring about its obsolescence.

This quick sketch of the emotional reception is quite unfair to almost everyone, but it helps convey some of the irritation and bias that is often mixed with enthusiasm when a professional group is faced with a new set of concepts. The actual reactions are far more complex than can ever be described succinctly. One needs for this task the verbal precision and the multiplicity of brief impressions achieved by poets, but there is hardly a poet that lives who will address himself to the subject.

The special character of the impact is also revealed if diffusion across academic boundaries is traced. The dynamic growth phase of communications theory originated in the Bell Telephone Laboratories and entered a period of tremendous extension immediately thereafter at M.I.T. and Harvard. In Cambridge between '48 and '51 the intellectual usefulness of these concepts was one of the most exciting features of academic gossip, but since then the intensity seems to have lessened. Circles have developed in other academic foci across the United States and have also started to spread eastward across Europe, even into Israel. Physiologists and psychologists studying networks of nerves were among the first to accept the new contributions, perhaps because they were already applying cybernetics concepts. Quickly the basic ideas were applied also to aspects of human communication-perception, learning theory, and linguistics. Physicists tied it to quantum mechanics in 1951 thus completing a full

triggered off, and that the reverberations from this explosion will continue to be felt for some time. Very likely no one is in a position to reflect upon it with an historian's perspective. My own survey has been conducted at a point somewhat away from the center of things (the United Kingdom and the University of Chicago); it has been carried out by an outsider who has always been sympathetic and intensely curious, but still a nonparticipant depending upon published reports as the principal source of information. Such a role brings with it some characteristic sources of error, but it permits the achievement of some perspective in judging long-range consequences. This review will therefore be an attempt to apply a wide variety of ways of thinking in order to gain an understanding of the social and economic effects and to state as briefly as possible the most obvious conclusions.

circle, because by virtue of hindsight we can now see that the first visible evidence of this mode of thinking is to be found in a paper by Szilard (5), stimulated when quantum mechanics itself was a new and exciting breakthrough two decades earlier. Library science was started off in a new direction. There were scattered papers, even books, in anthropology, philosophy, political science, sociology, and other disciplines. It appears that biological theory, from genetics to the functioning and the structure of the organism, will be fundamentally affected. Of all the established academic disciplines only economics is so isolated as to remain virtually unaffected to date (but there are rumblings from the research centers which suggest that in the long run it too will become profoundly involved).

Many scientists have misjudged the mode of application of the theory. They have sought to solve some of the most puzzling and controversial problems facing their discipline, which might be learning theory, political systems, or the structure of the chromosomes. This was to be achieved by translating Shannon's theory and its extensions into the language of their own discipline with the intent of discovering in the theoretical structure some hitherto undiscernible hiatus or inconsistency. This might be called the direct, or frontal, approach. It has been notably unsuccessful, perhaps because in any thoroughly argued scientific subject the theorizing has already been carried as far as it can go. The unique contribution of the new communications is to be found more at the periphery of the disciplines. It provides principles for translating the units for one system into the units of another. Hitherto, for instance, we have had one set of languages for describing interactions among machines, another set for interactions among humans, while the machine-human interaction was clumsily described in both kinds of languages, but without rules for transformation. Now we are provided with a system which tells much better than before whether the greatest efficiency is being extracted from the machine-human interaction. It is this property of tying together the information-gathering possibilities of many disciplines which offers the greatest promise. If the base is broadened, the pyramid might be built higher. Such an approach is more indirect because it means working at the periphery of a discipline before attempting to hew a new capstone.

The really important feature of the new communications theory is that it appears to deal in a quantitative manner with the flow and storage of information of any kind (some philosophers feel that there may even be a way of dealing with the meaningful, or semantic, aspect of a message). More than that, enough elaborations have been made already to suggest that there is here the potential for a still more general theory which will possibly be a *theory of organization*. One feels in reading current articles a common undercurrent of expectancy. The field is ripe for someone with extraordinary genius who will state the fundamental proposition. When, where, and how it will occur cannot be predicted, but it does appear that, for the world of ideas at least, the best is yet to come.

Changing the technology

Radical ideas have a peculiarly uneven impact upon contemporary practice. Those organizations which have already been undergoing change—sweeping away anachronisms like cobwebs from dark corners and modernizing their course of action as is expected of self-admitted pacemakers—often receive an extra push. They may move faster in a direction in which they have already achieved momentum, or they may take up a new direction. In either case they are likely to unleash a series of technical innovations and exhibit a spurt of expansion.

A few of these organizations, whose records for progressivism and adaptability are as good as any, become engaged in a great internal debate. Is this set of new ideas a temporary fad or a piece of pure charlatanry or, as some will assert, the forerunner of future methods for handling operations? The executives may ultimately decide to ignore the ideas altogether, or they may delay in deciding so long that only second-rate and third-rate opportunities remain.

There are quite a few other organizations, tradition-bound, already creaking at the joints, which are oblivious to revelations in the world of ideas. They tend to dwindle away into obscurity or lose their identity in consolidations.

And then there are many small, new firms, (also departments, bureaus, institutes, offices, centers, etc.) which are organized to exploit a vision. In them the mortality will be greatest of all, mainly because their resources are so limited. Those that do survive will be quite unusual organizations which are more truly geared to modern times than any other element in the society—so much so that they become a virtual parody of the contemporary patterns.

In the procession of names of organizations and their leaders, of policies announced, modified, and rejected, the channels through which the ideas flowed are not apparent. Yet it is by understanding the channels, their capacities, their rates of flow, their directions, and their biases, that we may estimate in advance the impact of a set of new ideas. This quest takes us backstage where the discussions the design engineers have with each other, and the arguments they have with management, may be intercepted. The engineers began reading about remarkable contributions to communication theory five to eight years ago. What have they been doing with them?

The first reaction has been very often an esthetic one. Some of the most competent engineers see beauty in a mechanism which works without human intervention, or a process which actuates its own controls. They see the theory offering neat solutions to problems which had hitherto always been messy. Communications theory for them provided a powerful analytical instrument which enables the composition of new symphonies of dynamics. Often it was possible to come upon a debate between proponents of the digital approach, which is based ultimately on the elements of logic, and the adherents of the analog approach, which permits the use of smooth curvilinear relationships, where all the arguments flowed from deeply felt ideals of symmetry and order. The issues were settled no more satisfactorily than would be a debate between Bach addicts and Wagner *aficionados*. In such an environment the fully automatic solution is often still too gross, so a fully electronic equivalent may be striven for with all the purposefulness and ardor of the creative artist.

Most competent engineers refuse to be dominated by this urge for internal unity, but their thoughts are nevertheless pushed into the same direction by demands from "higher up" or from "outside" for higher quality in the product or service for which they are responsible. They know from experience that automatic controls, suitably linked to each other, frequently lead to better quality control, or allow higher standards of performance in the product, or both. In order to reduce the frequency of errors in any production activity, the engineer finds that each step in the sequence must maintain communications with every other step and that workers, with their individuality and curiosity, are often major sources of errors. Thus even the more phlegmatic members of the profession found that their solutions would adopt fully automatic principles wherever they are indubitably feasible. Through their efforts in the past few years thousands of semi-automatic processes have become, or are due to become, fully automatic, while only one or two have gone back to manual techniques.[2]

However this trend has almost certainly been over-publicized in the past year or two. *Automation* has become a symbol for scaremongers, and copywriters for the national weeklies have stretched its meaning beyond recognition. Too often it is forgotten that technical innovation has difficulties of its own. To the man given the task of invention it seems that virtually all the ideas which are easy to generate have already been produced, usually in someone else's laboratory. Most of the worthwhile objectives require the prior accumulation of large

[2] These processes are major steps or stages in a production setup. James R. Bright has provided a very useful appraisal of the degree of automaticity in a recent article in the *Harvard Business Review* (1).

amounts of data, coupled with systematic borrowing from analogous situations. Occasionally something quite new must be originated. There has been no explosive increase in the engineers' problem-solving capacities—but there has been a remarkable increase in scope.

The current shortage of engineers is another reason for believing that the technology cannot change as rapidly as prognosticated. There are perhaps no more than 10,000 applied scientists and engineers who could fruitfully employ the communications principles in their work. More than half of them are in the hire of one or another defense agency, so that their most ingenious ideas are imprisoned in classified documents. The total is probably not increasing because the output of the universities is currently at a very low level. Therefore government departments, and the various industrial organizations, can expect to be short-handed for at least a decade to come, and very likely more. Many projects which would appear to be worthwhile for the long run must be postponed or laid aside altogether. Much of the social change predestined by the discoveries must wait.

One major conclusion can be drawn from observing the changes which have been set into motion. *The cost of information transmission, data processing, and social communications—considering the field as a whole—should decrease continuously for at least a generation.* Compare this prospect with that for other inputs, such as labor, energy, water, food, and mineral ores where the costs are moving strongly upwards, or transportation and capital where the trend is mixed. If a revolution is upon us, it is this fact which will be at the bottom of it all. A whole stream of consequences may be deduced as following upon such a trend.

Everywhere in society people will be considering whether it is not possible, in the interests of efficiency, to use one or another of the new and more elaborate means of communication to substitute for some standard ingredient which has become increasingly expensive. Thus teletype, and then ultra-facsimile reproduction, through microwave relays, might replace tons of paper, a battery of secretaries and typists, and a great deal of dependence upon the postal services. Later on, interdivisional conferences may be held by using leased wire television, thus saving many hours of valuable time lost in travel and a lot of transport costs.

The efficiency engineer's war against waste is aided at many points by introducing higher quality products which fulfill more stringent specifications. More information and over-all knowledge about the performance desired usually results in a demand for a more carefully tailored product. Control of the operations required to produce the better product also requires more information. So one sees, after much guarded talk between the two, a better informed producer linked with a better informed consumer. The net effect is an over-all gain in information, the improvement of product quality and a reduction in waste. Therefore the continuous improvement in quality, which was observed earlier as being perhaps the dominant practical reason for the engineer's acceptance of communications theory, seems to be just another facet of a general trend. It is a more intricate means for substituting information storage and communication for other inputs.

There are some rather obvious consequences that follow from these observations which will be stated here very briefly. The electronics and allied industries, including instruments and components, which has already grown to exceed $6.5 billions, seem destined to become the largest industry group in North America. The physical volume will increase much faster than value of output because the prices charged for assembled units are steadily decreasing. The number of production workers used may increase for a while but then should level off, because automation is moving more quickly into this industry than most. Maintenance and repair of the electronic and related equipment however should remain a burgeoning activity, employing millions of extra workers.

Along with the expansion of industry we may visualize communications invading every part of the society. Miners, foresters,

farmers, cattle-raisers, for example, have traditionally worked in surroundings which made communications difficult and expensive. Their jobs require too many varied judgments for an automaton to handle economically. They are thus *not* likely to be displaced by the fully automatic process. Nevertheless their work can be made much more productive if they keep in constant touch with developments in weather, the markets, the availability of equipment, the location of ore bodies, etc. Though the forester, farmer, or miner of the future may be no more conscious of it than he is today, he would be participating in scores of communications networks, a cog in at least as many organizations. The lonely life will be hard to find.

It is impossible even for scholars to visualize how each institution is likely to adjust as a result of the decreasing cost of communications. In some instances no amount of study seems to result in a reasonably persuasive picture, but quite a few inquiries do lead to a view of the future which fits all the trends. Most are in highly specialized industries, but in a few—like that of urban traffic—the public has quite intimate knowledge concerning the present state of affairs. Let us see, then what communications can do for the transportation problem.

An example: traffic regulation

In the past year or so some moderately elaborate computers have been made available for the control of stoplights. The signal is changed according to the numbers of cars approaching from the respective directions. Rules are normally incbrporated in the computer which will cause it to bunch the traffic into pulses or platoons, because a network of streets works best when vehicles move across it in small bunches. But sometimes this system of having each individual stoplight optimize flow through its intersection will lead to a horrible jam. The prevention of massive traffic tie-ups requires collecting traffic flow information at key points and feeding it into another form of computer, of a kind now being developed to handle aircraft, which can predict the potential jam some minutes or hours ahead of time and suggest measures for preventing it.

Long before this master computer was in routine operation it would be realized that traffic engineers had far too little information available to them. A computer and feedback system may minimize delay time for drivers and their passengers, but is it worth the expense? How much is delay time worth, at different times and in different places? Then, too, there is the closely related parking problem. What is the economic value of parking space under these various circumstances? And how is accident prevention to be worked into the scheme? Obviously many kinds of data would have to be collected continuously and fed into the new computers. There would have to be information about the choices drivers make when marketing or churchgoing, when the Yankees come to play a double-header, when it rains or snows, and whenever there is a popular TV show.

Inevitably, at various points in such a network, traffic flows will grow through the years to such a level that the computers cannot cope with them. Losses due to delay time start spiraling upward. A major physical alteration of the system becomes necessary. One solution to the difficulty may be obtained by pouring hundreds of millions of dollars into new freeways. However calculations would show that it is cheaper by far to build mass transit—even if the service were free! And it would pay the city to use gasoline tax money to do it—if that remained the only major source of income for effecting large-scale changes. Then, it would quickly become apparent, new kinds of data would be required which are not vehicle oriented but shipment and passenger oriented—which means more data collection points, larger communications networks, bulkier statistics, and bigger computers, because then the interchanges between several different systems of movement (airplanes, helicopters, railways, buses, passenger cars, etc.) would have to be continuously optimized as well. Very likely fully electronic systems for data processing would be necessary at this stage. Only the slow sub-

stations and the farther satellites of the metropolitan district could afford to fiddle with paper tape and punched cards.

At that time the foundation would have been laid for quite new approaches. The new voice-powered transistor radios (Dick Tracy type) should be available for less than $100. Wouldn't it be much cheaper to employ a system of taxi and car-hire than to buy a second and third car for the convenience of the family? Persons with handy wrist-size radios, wishing to move from one spot to another in suburban wilderness, would not need a private telephone to request a car from the local pool. A great deal of the saving would come from the more intensive use of such cars (say 30,000 miles per year instead of 6000). Millions of parking places would also be saved, and cities could become much tidier in their built-up parts. In this instance an expansion in the communications system would take the place of a much more costly expansion of the highway system, the supply of automobiles, land allocated to parking, etc. Perhaps in a decade or so the anti-collision radar equipment for autos will also be ready for introduction. Dozens of other communications-based schemes like this one can be devised, some of which will, no doubt, be used widely.

The consumer would not see many radical changes that could be attributed to the breakthrough in communications science. Virtual invisibility is one of the peculiar features of this technology. Even the newly acquired housetop vista in the suburbs, where TV and FM antenna have sprouted in chaotic profusion, is likely to pass. The architect, the builder, and the planner are faced with new specifications for structural form and siting of buildings which assure good service in the ultra high frequencies (UHF), but these requirements hardly seem likely to call into being new architectural styles, or new layouts for cities, all by themselves, as has happened with the automobile and electric power.

The traveler would note that fares, tolls, rentals and other charges were not rigid but actually highly variable according to time and place, being especially steep during rush hours, and that residents found it more convenient to use credit cards and be billed for transport services much the same way oil companies do today. Accidents would be rarer, and streets would be cleaner, but not strikingly so.

However if ever the consumer should be so curious as to investigate the system which had evolved, he would discover that a gigantic market had been set up so as to adjust the supply of services to meet the demand as soon as it was evidenced. This is a place where price theory in classical economics, even in its purest form, can be put to work. Large metropolitan areas like New York, Philadelphia, Detroit, Chicago and Los Angeles would find it to their net advantage to invest billions of dollars into communications equipment which coordinated movements if, by such means, they saved only ten to twenty minutes per person per day. And the chances are quite good they will have to spend those amounts—or stifle.

The advances in communications make it possible for American society to evade the full effects of the mistakes in urban development which have already been made. They permit economic growth in new and unexpected directions.

THE SUBTLE EFFECTS

The significant consequences of the breakthrough in communications theory are not to be found in the rapid growth of the electronics industry or the changes in structure in the labor force. They go much deeper than that. A new way of life, the dimensions and significance of which cannot be grasped as yet, is opening up.

The cheapening of communications equipment, per unit of physical throughput, leads to many modifications of behavior. We see the process at work every day in our own environment when we ponder how much more dependent upon telephones, radios, television sets, and recording devices all the people around us have become. The changes are still more apparent when it is observed how a corporation, government bureau, or a school, or any other conservative institution, must change in order to stay modern and up-to-date. It is most

striking of all to see how certain crucial organizations like banks, libraries, and traffic systems would have to revamp themselves quite thoroughly if they are to fulfill their functions satisfactorily. Such changes in organizational behavior can be deduced, one by one, some more clearly than others, but it is more difficult to represent the modifications and additions to be expected in the *way of life*.

Nevertheless an introduction to the new patterns of living appears possible. It may be based on the changes in communications behavior which seem at the moment to be most probable. They take the form of ten propositions. From them one can make quite a few deductions about social change, but to fill in the picture each person must become his own prophet. The propositions are set forth here mainly to discover whether there is not some mutually agreeable starting point from which each person, with his specialized experience, may judge the consequences of this advance in science and technology.

1. Communications are transcending practically every physical barrier, including distance, altitude, heat, darkness, crows, the aurora borealis, the lack of electric power production, and many others. There is a communications system available which is suited to almost every time and place, and most of them are destined to become less expensive as time passes.

2. Crucial failures in point-to-point communication, which show up so frequently in the study of historic decisions, are much less probable today. They are still less likely in the future. Mistaken decisions in the future will be due almost entirely to poor planning and to imperfect social institutions, not to the physical inadequacy of communications systems.

3. The volume of communications is growing at a tremendous rate. A very tiny amount of this is crucial to the orderly functioning of human affairs. The remainder deals either with low priority interaction and details or with the development of leisure time activities such as the arts, sports, or hobbies.

4. If categories of communication are set up which distinguish between

a. face-to-face conferences, meetings, gossip, etc.,
b. reading newspapers, magazines, books, billboards, etc.,
c. man-machine, reading instruments, gauges, microscopes, radar, etc.,
d. person-person (*machine interposed*) telephone, radio, TV films,
e. machine-machine, feedbacks, interlocks, automatism,

it becomes possible to gauge specific rates of growth in communications. There are grave difficulties in establishing equivalences of communications flow, which would very likely require extensive investigations before satisfactory measurements and comparisons could be made, nevertheless one can already achieve a feeling for the crude relationships anticipated. These are shown in Figure 1.

5. Face-to-face communications have been increasing because of urbanization, improved transport, and more carefully integrated social organization. They are expected to decline somewhat in the future because other methods of communication (mainly machine-interposed) will serve as substitutes.

6. Reading has been advanced by a reduction in cost of publishing relative to income and also by the contemporary needs of social institutions where virtually all reports, records, and documents are written. It is expected to decline later because of substitutions from the following two categories.

7. Man-machine communications are expected to rise slowly. They are at present intended mainly to acquire useful information from the environment which is beyond direct contact with human senses but later they will constitute reports from machines of the accomplishment and failures of the machines themselves.

8. Machine-interposed social communications are expanding most rapidly at the moment owing to television and related equipment. It must level off sometime owing to saturation of human perceptions. This may be 10–100 times the present meaningful communications input for the average individual. Further improvements are possible

but only through drastic revision of the languages, symbols, and codes employed in the society.

9. Machine-machine communications are used primarily now for adjusting to variations in temperature, pressure, composition of input or output, etc. In the future they will tend to deal more with interrelations between machines in a sequence of operations. As explained earlier there seem to be no ultimate limits upon this form of communications except cost, and costs are decreasing. Therefore its growth has the properties of an explosion.

10. The content carried by any communications system can modify people's perfunctory day-to-day decisions. The content from several *apparently independent and reliable* channels and sources over time may induce them to change their convictions. Thus these systems taken together are an instrument of power for institutions and organizations controlling content. They are also a means by which individuals may more successfully adapt their actions to their social environment.

It will be seen that the flow of significant information, on a per capita basis, is likely to be multiplied many times over in the next generation. For anyone who has observed machines with electronic circuits, and understands the parallelism with the nervous system, this introduces some disturbing new problems. Perhaps the most important of these is the question of "communications-saturation." Is there not some point beyond which the information can no longer be assimilated? At the present rates of increase in communication flow it seems quite possible that whole sectors of the society may be facing saturation in the not-too-distant future.

In machines and communications networks, the approach to saturation is asso-

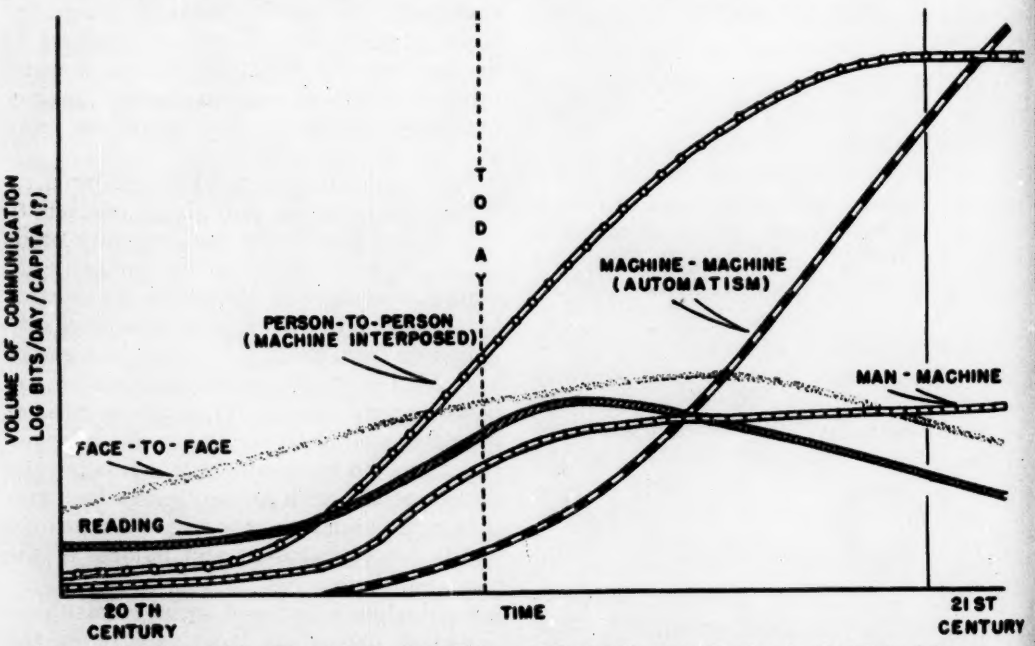


FIG. 1. Relationships Between Various Forms of Social Communication in the United States

Automatism has quite different properties of growth than the other forms of communications. The relationships shown here are based upon observation and therefore represent only the abstract properties of each form of communication—their trends and their relative dimensions. Perhaps in a decade or so we may be able to reduce these to a common denominator, an encoding device, which would permit their actual measurement. The flow of information in a society has many of the properties of the flow of money, so it is not unlikely that many of the familiar principles of economics will be extended to social communications as well.

ciated with the jamming of messages, greater frequency of simple errors being transmitted, and the mixing of addresses and instructions. Sometimes the system becomes unstable so that it alternately races and dawdles, with the condition becoming progressively worse over time so that it ultimately breaks down. These properties in machines are suggestive, but human personalities and organizations are far more complex and thus likely to have quite a few more stability points, many of which could accommodate larger flows of information. Very little is known as yet about the effect of greatly increased communications within society, but certain simple experiments have been made and some relevant observations are possible.

Perhaps the most interesting studies of this kind were carried out by Quastler and co-workers at the University of Illinois (2). They sought to find the maximum throughput of information by humans. A very good performance was achieved by skilled pianists who read musical notes at a top rate of about 22 bits per second. This throughput is about the same as that which can be achieved by humans who have a knack for rapid arithmetical calculation and it is about what can be reached in silent reading (say 25 bits).

The task of playing written notes of music on a piano may be considered to be merely that of translating one set of symbols into another (written notes into keyboard fingering) where the decision rules are direct, unambiguous, and uncomplicated. Inability to recall what action corresponded to a given symbol was ruled out by using professional pianists. The use of experts however introduced a new form of bias based upon their experience. Certain clusters of notes may reoccur as a group, in much the same manner that words may be used in a cliche! Therefore several passages of random notes were constructed for the experiment. When the subjects were pushed to capacity (by an accelerating metronome) they appeared to develop quite different strategies in response to the stresses of communications saturation. One man seemed to emphasize speed of response, letting accuracy suffer, while another tried to minimize errors by letting speed drop off in the hope of reaching an "easy" passage which would enable him to catch up again. A third seemed to be attempting the more complex task of balancing increased errors against losses in speed.

What happens when memory is allowed to play an important role? How much is transmitted then? The scientists turned to their own behavior as a clue, using the recording tapes as before. An organized, but extemporaneous, lecture could deliver up to 18 bits per second. (These are still very preliminary estimates because, in order to obtain any comparable figures at all, a number of simplifying assumptions had to be made. As the experiments become more sophisticated the data become more trustworthy.) Scientists can attest to the fact that this pace is exhausting. Two consecutive hours of it leave one quite limp, and five hours a day at spaced intervals is about the highest rate we have seen hard-driving colleagues maintain.

Other clues to the effects of communications saturation may be found in the communications businesses themselves (radio, television, cinema, advertising, education, etc.) and elsewhere in certain foci of communications where one normally finds executives, administrators, consultants, professors, doctors, and the like. Such people live in an environment which, because of the volume of communications to which they are exposed, forces them to work out systems for scanning, selection, and criticism of the inflow. Their techniques can become elaborate. One method might be that of establishing a very thorough filing system, carefully cross-indexed, so that the history of the category or case, or a close parallel to it, would largely determine the action to be taken in the light of the newest information made available. Other persons in this position, especially those who bear a heavy responsibility, will start delegating the information-handling tasks. The input may be analyzed by a team of specialists, one to analyze cyclical properties of the statistics, another to discover those situations for which standing orders exist, another who would look at the long-range aspects and take

care of the planning, etc. On the output side one finds the inevitable public relations man, perhaps a labor relations man, and executive assistants whose function is often that of carrying complicated messages to other executives and administrators. Thus one or more teams may be organized to process communications on their way to and their way from the decision-making focus. Their task is that of distilling from the flood of material the essential and relevant items which can be digested by one man, or a committee, and maintaining mechanisms by which a relatively few words of decision can be amplified into action on a broad front.

The communications-saturated individuals in this environment are, very often, alert and sensitive persons, subject to stomach ulcers and cynicism, but still active participants in society at large. In off-hours one sees them heavily involved in professional and community organizations, frequently a great variety of them, and they often lend their sympathy, expressed in the form of checks through the mail, to many other causes about which they remain reasonably well informed. All this makes for a lively and wearing life, and one which the population as a whole probably could not tolerate.

A common reaction to being exposed to high volume communication for some time is the search for some authorized excuses for occasionally cutting all connections in order to relieve some of the pressures. Thus for many such persons holidays and vacations become sacrosanct. They need the chance to relax, reflect, be entertained, and in general behave spontaneously as individuals rather than as functionaries with carefully specified roles. Some will take time to break the rhythm of reading and writing, of seeing and expressing, by working with their hands, painting, or hiking in the wilderness. Some of the old, experienced hands at this kind of life develop a habit of rising at dawn so that they have at least an hour to reflect on matters in general before the newspapers are picked up, before the telephones dare ring, and before the eternal game of passing messages shifts into high gear.

Changes in social institutions

The most fundamental reaction to increasing information flow in society is the adjustment of each of the separate organizations so as to cope with it. There is no theory to be drawn upon when making these adjustments, so each executive and each functionary is forced to experiment. He draws upon his intuition and experience and proceeds to improvise. Yet there are several generalizations about the adjustments which can be made—perhaps because of the commonality of the "common sense" which is used. It is worthwhile summing up these observations because they may help someone construct a satisfactory theory.

One of the first steps is to clarify the goals and objectives of the organization. (In the words of younger businessmen this takes the form of asking the question "What business are we *really* in?" If this question can be answered and agreed upon, then a hierarchy of goals and objectives can be quickly deduced from the known structure of the economy.) Once these are known the values or weights (degrees of importance) to be applied to incoming information become virtually automatic. Rules can be formulated so that the sifting can be carried out satisfactorily by an assistant to the decision maker rather than himself. The least important information can be discarded at this stage. A subsequent step is to attach an address to various categories of information so that it is relayed to the group (or groups) in the organization whose task is affected. At the same time the channels for the flow of information within the organization, whether upwards, downwards, or laterally, must be expanded. This will mean adding to overhead expense, first for typists and secretaries, then for machines, like office intercom systems, dictation recorders, quick reproduction equipment and similar items, to help deal with the volume, and then for a multiplicity of special services being made available in the business world outside of the organization. It may end up with a thorough revision of the methods of drafting and handling internal reports.

One rather obvious set of measures will be

taken within such organizations. The equivalent of noise within the internal communications network (irrelevant interruptions or disturbances) would be reduced. This permits a reduction in the amount of redundancy (repetition) within the messages that are being relayed back and forth.

There are also special emotional constraints which operate to filter or choke off the flow of significant information. Some of these are due to the traditional status relationships in the organization, as when the underling is afraid to tell his boss the kind of truths which will offend him. Wherever biases are introduced due to intense emotions (like that engendered by status difference, sexual relationships, religious feelings, ideological ties, etc.), there are measures prepared for safeguarding the communications system. These may replace the offending symbols with equally good ones which are emotionally neutral, or certain personalities capable of blocking the more crucial messages may be excluded from the network. The sources of emotionality need to be known and communicated, but it is more convenient to segregate them into separate channels.

An organization might be expected to contain an interlocking assortment of task groups, each with its own subgoals, skills, resources, and decision rules. (A decision rule is a principle for action. If certain specified conditions prevail, then the experience says that one particular action brings better results than any other. In baseball, for instance, if the score is close, nobody is out, and a man is on first base, the most common decision rule calls for the batter to bunt, sacrificing the runner to second. As in baseball, any given group may develop hundreds of decision rules, many of them based upon decades, or even generations, of experience, which will cover most situations encountered and save thinking and argument at the time of decision.) When the flow of communications increases steadily, the repertoire of decision rules must be reorganized. Many more identifiable situations need to be covered with new decision rules at the same time that a portion of those on hand become obsolete. One might even begin to think of the normal lifetime of a decision rule under these circumstances, so vulnerable are they to change. However the decision rules for the various task groups need to mesh so that the aims of the whole organization are accomplished. Nowadays a special planning unit is often created to study continuously the long-range problems and recommend changes in decision rules and the further addition of task groups. In many corporations it is called O & P—organization and procedures—while in government several new agencies and commissions share this function.

Changes in social values

Let us transfer our mental images of communications-saturation to the bulk of the population several decades hence. Through communication each person, with his actions and his goals, will enter into the life of many other persons far more than has ever been true in the past. The barriers in communication which up till now permitted each ethnic group and social circle, even those in the cities, to create a full life of its own, will be overwhelmed by the mass media and the demands of large organizations. People will have much more experience in common, not in the form of cultural tradition because it becomes archaic and laughable too soon in a time of rapid social change, but in the form of a standardized education, up-to-the-minute appraisals of the social, economic, and political environment, technology and science.[3] The potential for helping and hurting each other in achieving individual aspirations is increasing every year. (It will be recalled that *automation* in industry and commerce leads necessarily to the employment of more persons in the service jobs which attend not only to the machines but

[3] This is most evident in the hard thinking now going on concerning the types of programs suited to educational television. A colleague, Reuel Denney, in a recent highly insightful address elaborated the kinds of innovations such programming would have to make if it were to be successful. Each of these would lead to substantial increases in the flow of significant communications; each is a long step toward saturating society.

even more to the persons responsible for maintaining the machines.) Therefore, out of these developments which bring members of a society into much closer and more intimate contact, particularly the individual with the group or mass, a finer sense of propriety must evolve. We will need much better means for establishing whether intrusion is an offense against privacy or a welcome release from inactivity and loneliness. The new rules must be worked out by wise men and philosophers, and tested by experiment, because the ethics of social contact handed down by our parents and teachers is already inadequate to the situations thrust upon us by the new communications media. Witness the conflicts aroused by the singing commercials as an apparently logical step going beyond the advertising displays normally presented in buses and trolleys!

The synthesis of new rules for behavior will be affected by the opportunities afforded for creative activity. Rules bring with them, if they are wisely formulated, both constraint and freedom, but freedom in a larger measure. The freedom is sure to be abused on some occasions, because people everywhere find means for playing with new tools in as simple-minded a manner as they have toyed with the artifacts that preceded them. A communications channel can be used frivolously so that much time and patience are wasted, but out of this, it is hoped, will come a sense of proportions, and a feeling for the fundamental properties of the new medium. Gradually it may become an outlet for disciplined creativity and novel forms of art can emerge. There are evidences of this in the American society, which has pioneered in opening up network television, FM, tape recordings, high fidelity transmission, adioautography, and others. Americans have the opportunity, because of wealth and prior discovery, of being the first society to master new modes of esthetic expression, not only within the media mentioned, but in others yet to be created. Understanding the new communications theory should lead to cross-fertilization between media and a realization of the scope for synthesis. The theory, and the gadgets derived from its principles, should have the same relationship to the artist in the new media as the musical instrument makers—backstopped by researches in acoustics—have had to the composer and virtuoso, and the camera technicians have had in the making of art films.

The preceding sentences were written some months before the announcement of the electronic music synthesizer by RCA (3). It permits the translation of a composer's score into the sound that would have been recorded if the respective instruments were played. The results so far achieved, after a few years of effort on the part of a small team, are at a kind of half-way station. What comes through on the recordings is a faithful reproduction of the notes—the composer's contribution to music—but none of the personalized impressions supplied by players of instruments. It is competent music, but it is not alive. A scientist would say that the fine structure was missing. However these things can, through experiment, be incorporated. Then there would be available not only a wide variety of timbres, rhythms, and harmonies which have never been physically possible before, but the possibility for a single composer-artist imbuing a full orchestral composition with his own feelings and susceptibilities—without fussing with conductors and musicians about their interpretation of passages. The mathematically inclined see for the first time ways of summing up the disciplined impress of a master musician and so it is not unlikely that they may be able to define characteristics of new schools of interpretation which, perhaps due to chance alone, have not been established in the course of musical evolution. Then too, as every hi-fi enthusiast already knows, there is an immense region of synthesized sound for which notes cannot be written, but profound sensual impact is nevertheless evident. It is in this area that nonclassical music will find its happy hunting ground.[4]

Very similar innovations are possible with visual phenomena, only—according to com-

[4] The first public recognition I have come across regarding the potentials of some of these channels is Roger Moren's review, "Music by Montage and Mixing" (*Reporter*, October 6, 1955, p. 38) of the developments based upon the achievements already attained employing tape recorders.

munications theory—the syntheses are likely to be even more complex. There seem to be a greater variety of distinctions which can be drawn and a larger number of levels of abstraction inherent in sight, as against hearing or touching or tasting. To say that one should expect abstract art in color and the full four dimensions is still an understatement—it must deftly weave together several levels of abstraction. The languages of vision are still so imperfectly understood!

However, one matter stands in the way of all such developments. There are as yet no institutions where competent and creative people can come together under relatively ideal conditions and experiment in such a way that experience rapidly accumulates. Universities have never really taken in many creative artists, although they occasionally strike up an interest in such an extension of their activities. The faculties are quickly frightened away by the discovery that the equipment for this activity must be as extensive as that for a physics department, and there is no military budget to share the burden in the support of the arts. Colleges usually compromise by putting a poet or a composer on the payroll, since they are still cheap. But if we are to fill the new channels with something more than drivel and threadbare stereotypes, and not wear out the old masters of music and drama by too frequent transposition and repetition, then a much more organized aggregation of creative talents is required.

Universities of the modern arts with creators, critics, and students, endowed with equipment, a measure of security, and a social responsibility, dedicated to experiment with abstract, complex symbols, and the production of new works of art . . . is this not a logical necessity springing from the communications development? Institutions which exploit the freedoms inherent in communications are rather crucial, because otherwise any straightforward extrapolation of modern social trends to which these new technologies will be added leads the scholar to the contemplation of a future as dismal as that of *Brave New World* or even *1984*. The techniques for using multiple channels for propaganda as a means of exerting political and economic influence are quite well understood already, so that improving the technical facilities at the disposal of authority makes possible the substitution of manipulation and persuasion for outright compulsion. Masses of people can be duped into behaving in a fashion convenient to the cliques controlling the various communications media. The mounting horror induced by the exposition of these "wrong way" utopias, where technology becomes a spectacular instrument of totalitarian oppression, may be attributed to Huxley's and Orwell's inability, or unwillingness, to understand that the evolution of new modes of expression, interacts with the capacity to accept subordination. Under most conditions an individual obtaining more information than he previously had, also acquires greater freedom; he has a wider set of alternatives to choose from. He has more viewpoints at hand for comprehending a problem. No one can yet envisage the changes in the conception of what is worthwhile, or the readjustment of personal values, which should occur in the light of the new opportunities. These are affairs that must be worked out as we proceed.

What is being suggested here is a strategy as old as history. A new territory is being opened up for exploitation. It is not virgin land on a newly discovered continent, or a new classification of products and processes in the patent office, but the appropriate strategy appears to be much the same as was employed in these areas. If this new territory is explored and developed initially by persons or institutions with private (in this case esthetic) ends, rather than for purposes of power and socio-political control, the "vested interests" which come into being will reflect these same private values. It is not certain that totalitarian manipulation would be permanently excluded but prior occupation by a variety of neutral and antithetic interests would make its intrusion more difficult, and therefore less likely. Even in the course of succumbing, if that were inevitable, the heirs of the early settlers could exact many compromises. Those who wish to extend meas-

urably the freedom of individuals and small groups to engage in creative activity must establish their claims in the next few years.

The point is that, if we neglect the new frontiers in human expression and cultural growth, the cheerless prophecies of political decline, moral decay and doom all have a much better chance of fulfillment.

REFERENCES

1. Bright, J. R. How to evaluate automation. *Harvard Bus. Rev.*, 1955, 33, 101–111.
2. Quastler, H. and associates. Human performance in information transmission. Control Systems Laboratory, 1955, No. R-62.
3. Sarnoff, D. New developments in electronics. *Elec. Engng*, 1955, 74, 181–185.
4. Shannon, C. E. A mathematical theory of communication. *Bell Syst. tech. J.*, 1948, 27, 379–423, 623–656.
5. Szilard, L. Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen, *Z. Physik*, 1929, 53, 840–856.

(Manuscript received October 24, 1955.)

Most philosophers . . . have exhibited a grotesque ignorance of man's life, and have built up systems that are elaborate and imposing, but quite unrelated to actual human affairs. They have almost consistently neglected the actual process of thought and have set the mind off as something apart to be studied by itself. But no such mind, exempt from bodily processes, animal impulses, savage traditions, infantile impressions, conventional reactions, and traditional knowledge, ever existed, even in the case of the most abstract of metaphysicians. Kant entitled his great work **A Critique of Pure Reason.** But to the modern student of mind pure reason seems as mythical as the pure gold, transparent as glass, with which the celestial city is paved.

—James Harvey Robinson in **Am I Thinking?**

A CRITIQUE OF FREDERICK MOSTELLER AND ROBERT R. BUSH'S *STOCHASTIC MODELS FOR LEARNING*

by Anatol Rapoport

Mental Health Research Institute, University of Michigan

A stochastic process is one in which the probabilities of a set of events keep changing with time. Bush and Mosteller make use of the mathematical techniques developed for the study of such processes in building a theory of learning and then apply the theory to explain the results of several learning experiments. The reasoning behind their theory and its applications to one of the experiments are discussed here.

The approach to learning phenomena in *Stochastic Models for Learning*[1] is notably consistent in its orientation. As the title indicates, the aspects of learning behavior investigated are exclusively stochastic ones, and, as appears (sometimes explicitly) from the exposition, the method is exclusively behavioristic, except for occasional excursions into "pseudo-neurology" to be noted below.

Consistency is a virtue when it reflects not rigidity but a disciplined determination to pursue a method of investigation as far as seems feasible. This virtue the authors have achieved with the result that from the vantage point of their advanced position "at the end of the line" they catch a glimpse of new territory which bears a definite promise for future exploration.

The authors have accepted the limitations of the behaviorist view and have made the best of it: they have successfully axiomatized their assumptions, so as to permit a rather rigorous application of the stochastic method. Having defined learning as a systematic change in behavior (whether adaptive or not), they describe such changes in terms of relative frequencies of classes of responses. They assume the classes to be perfectly defined and distinguishable and further that a probability can be assigned to each class at any time when a response is being elicited (a trial). Besides these responses, there occur certain events which change the probabilities associated with the responses. The authors regard these events as outcomes of the responses, possibly inherent in the responses themselves, possibly instigated by an outside agency (such as the experimenter, who may mete out rewards or punishments and in other ways change the situation eliciting a response). At any rate, the outcomes are assumed to be as distinguishable as the response classes. The model is a set of mathematical assumptions concerning how the outcomes change the probabilities of the response classes.

This axiomatization establishes the approach as markedly behavioristic. The focus of attention is on outwardly observable identifiable classes of events. Except for a few departures, which are not essential to the main line of argument, there are no attempts to relate behavior either to neurological structure or to motivations, drives, perceptions, etc.

The mathematical treatment of the assumptions marks the approach as stochastic. Here perhaps some elucidation is in order for the technically uninitiated reader. One might remark that a chapter devoted to the peculiarities, advantages, and limitations of the stochastic method as compared with other quantitative approaches to the same phenomena would have helped the reader to

[1] *Stochastic Models for Learning* by Robert R. Bush and Frederick Mosteller. New York: John Wiley & Sons, 1955. Pp. xvi + 365, $9.00.

see the authors' effort in more vivid perspective. On the other hand, it was, of course, perfectly proper to let the stochastic method in action speak for itself, leaving the responsibility for comparative discussion and evaluation to the reader or the reviewers.

There have, in fact, been two other quite distinct quantitative approaches to learning phenomena, with which the authors acknowledge acquaintance. One, essentially an empirical-descriptive approach, is exemplified in the work of Hull (among others); the other, involving genuine mathematical (but not stochastic) models, in the work of Rashevsky, Landahl, and others. All three approaches use mathematical and statistical tools but differ sharply in their centers of emphasis.

The empirical-descriptive approach

The empirical approach always starts from data. Its hypotheses often go no further than asserting that regular quantitative relations exist among the variables it has chosen for observation. Thus in learning and conditioning experiments the empirical method may seek the relations between number of reinforcements of a conditioned stimulus and the strength of response, or between the interval separating the conditioned from the unconditioned stimulus and the rate of increase in the strength of response; between number of trials and number of errors in successive maze runs; between the assumed motivating drive (hunger, pain) and the rate of learning, etc.

The results of these inquiries are tabulated sets of data, which when plotted on graphs often suggest curves. The mathematical work is usually limited to a search for an expression for the best fitting curve, while the statistical work may be limited to the evaluation of the "goodness of fit" or else to a tabulation and comparison of the standard statistics of the data—means, modes, variances, etc.—with a view of evaluating the significance of the results.

The advantages of this method are its directness and its immediate contact with observed facts. The assertions that derive from it, such as, "curve $y = .037(1 - e^{-0.69x}) + 2.11$ fits the data with such and such accuracy" or "There is a significant difference at such and such level of confidence between the effects of two kinds of deprivation on the rate of learning," are incontrovertible except through discovery of experimental or calculational errors. However, in the strength of the method lies also its weakness. The absence of *a priori* theoretical formulation insures against commitment to a false conceptualization but usually delimits the opportunity for theoretical insight.

A comparison between the empirical-descriptive method and the authors' is not of great methodological importance, because the former does not purport to be theoretical, while the latter does, and so the methods really do not "compete." More illuminating would be comparison between the method of stochastic models and that of other mathematical models.

The stochastic approach

Any mathematical model, including the stochastic, proceeds from assumed relations among variables (which are often not observables but theoretical constructs) to *predictions* of quantitative relations among observables. The latter are deduced by often involved mathematical techniques from the former. Any mathematical model may arise from consideration of probabilities of events, but what distinguishes the stochastic model is that it keeps random variables at the center of attention throughout and in that the probabilities usually change with time. The difference between a probabilistic and more specifically a stochastic model may be best illustrated by an example from physics. The theory of the diffusion of gases is derived from probabilistic considerations involving individual molecules, but the end result is a set of equations describing *determinate* relations among certain variables (concentrations, space coordinates, and time). If, however, we were interested in the behavior of an individual molecule in the same situation, we would have to apply some stochastic model (e.g., the "random walk") in which the position of our molecule would at all times be a *random variable*, specified not by a value but by a distribution of probabilities over a range of possible values. This distribu-

tion of probabilities would change from moment to moment, and so our result would be expressed in terms of probabilities, just as our assumptions had been. Of course, a "probability" is never measurable directly: only frequencies are experimental indicators of probabilities. Thus even the stochastic model, if tested, implicitly predicts determinate relations. However, it is characteristic of a stochastic model that the probabilistic (indeterminate) situation is not lost sight of during the entire theoretical investigation.

Having made the distinction, we are ready to compare the stochastic approach to learning with other mathematical approaches. This we will do in the course of our discussion. For the present we turn our attention to the examination of the authors' model.

The model

Following the authors' example, we emphasize the simple case in which there are only two possible classes of responses. If, at a given trial, the probability of one class is p, the probability of the other is necessarily $1 - p$. Therefore there is in this simple example only one variable, p, on which certain operations will be performed. These operations determine how the value of p changes from trial to trial as a result of the pertinent events, which the authors call the outcomes of the trials. To each outcome, therefore, there corresponds an operation. The fundamental assumption is that the transformation of p is linear, as expressed in the following equation

$$Qp = Ap + B \qquad [1]$$

The left side indicates the operation upon p by an "operator" Q (not multiplication); the right side gives the new value of p, the result of the operation; A and B are two constants which define the particular operator Q. The generalization to the case of several responses is straightforward: p becomes a vector, whose components are the probabilities of the responses; A and B become matrices. Essentially, this is all there is to the model. The entire first part of the book (over half) is devoted to the investigation of its purely mathematical properties.

The question in the mind of the reader interested primarily in learning theory is inevitable: why *this* particular model? This question can be immediately answered by anyone who has worked with mathematical models: the simplest of all transformations is the linear transformation. Therefore if a model is to be based on a sequence of transformations of a variable, the first to be tried in the absence of any preconceived notions is the linear one. However, even a reader experienced in the use of mathematical models may well ask what there is about a linear transformation that takes 184 pages to discuss. The answer to this question is considerably more involved. It actually reveals the essential features, uses, and value of the stochastic approach.

Some of the mathematical properties of the model are quickly established. Thus it is easily seen that as a result of repeated applications of Q, p is either always increasing or always decreasing. Furthermore, Equation 1 can be written as

$$Qp = \alpha p + (1 - \alpha)\lambda, \qquad [2]$$

where α and λ are now the two parameters. Certain numerical restrictions on α and λ follow from the fact that always $0 \leqq p \leqq 1$. When written in the form of Equation 2, repeated applications of Q become

$$Q^n p = \alpha^n p + (1 - \alpha^n)\lambda, \qquad [3]$$

and since, in general $\alpha < 1$, p approaches λ asymptotically.

Different outcomes of responses are associated with different operators Q. In general, the operators are not "commutative", i.e., the application of two operators successively in different order gives different results. However, under special conditions the operators do commute. All these and similar matters are discussed in the introductory chapters.

The exploitation of the model

Results of more direct significance to the theory itself begin to appear in Chapter 4. Here the relation between the mathematical and the behavioral systems of axioms underlying the model becomes apparent. The key concept in the Bush-Mosteller theory is the

probability of a response (or of a class of responses). This concept underlies other approaches also but is not central to them. In earlier mathematical treatments a typical "learning curve" appears as a smooth curve drawn through a histogram of, say, errors plotted against trials. It is usual in a mathematical derivation of such a curve to equate its derivative to the probability of the "wrong" response. For the derivative is roughly the ratio of the increase in the accumulated errors to the increase in trials. However, this identification of "probability of error" with the derivative of the "learning curve" makes sense only if each interval considered contains a large number of trials. One cannot read off the "probability of error" from what happens on a particular trial any more than one can determine the probability of heads from a single throw of a coin. However, while the result of a single trial is frankly dispensed with in the theory of the learning curve, it occupies a central position in the Bush-Mosteller theory. Bush and Mosteller assume a large mob of "identical" individuals going through the sequence of trials. The fraction of these individuals making the response in question on a particular trial leads to one of the estimates (but by no means the only one) of the probability of the response.

In other words, Bush and Mosteller average over a population, whereas the usual procedure in the individual learning curve is to average over a time interval. On the face of it, both averages make sense only under Utopian conditions. Bush and Mosteller require large numbers of identical individuals (naturally many runs of the same individual will not do, because his experiences cannot be simply erased). A learning curve, on the other hand, requires the individual to be "constant" over a period of time, which, although short compared with the total duration of learning (to give meaning to the "derivative") is long enough to contain sufficiently many trials (to give empirical meaning to "probability"). However, and here is an important advantage of the stochastic approach, Bush and Mosteller are able to mobilize many other statistics besides the average rate of increase of errors to buttress the theoretical significance of their model. Indeed, the very simplicity of the model makes possible a most thorough application of statistical analysis *as part of the theory of the model itself*. The virtue of the model is "depth" rather than "breadth," as will appear from the comparative discussion below.

As has been pointed out, the theoretical analysis of the model occupies the entire first half of the book. This observation needs emphasis, because the purpose of such analysis is not too commonly understood. The actual construction of a mathematical model is only one phase of an attack on a problem by mathematical methods. In the case of the model under discussion, it was the most trivial phase, the model being the simplest possible and therefore the first that should have occurred to anyone concerned with the problem. The task of relating the model to data is another phase in the work. This phase is the most widely understood. But between the construction and the application there is an extremely important phase of pursuing the purely *logical* implications of the model. What can it be expected to reveal and what not (given a hypothetical set of data)? What sorts of data will give most information? How should the experiments be designed to get such data? These questions cannot be answered in the process of constructing the model, and it is too late or too costly to ask them after the experiment has been performed. They are the questions which must be answered in the intervening task of thorough logico-mathematical analysis.

The realization of the Bush-Mosteller model, like that of any stochastic process or typically like a "game in extensive form", is a branched tree. In the trivial case where there is only one operator (i.e., where the transformation of p is the same no matter what happens), the "tree" is just a series of steps. The value of p is determinate at each trial and proceeds to its limit. But if there is more than one operator (as, in general there is, e.g., when how p is transformed depends what happens at each trial), real branching occurs, because at each trial the

value of p may assume each of several values, depending on which operator is applied, and each of these again splits at the next trial, etc. In the simplest nontrivial case of two operators, p may have assumed at the end of n trials any of 2^n possible values. Moreover, there is a probability assigned to each of these 2^n values, determined by the probability of the particular sequence of operators that has led to it. There is thus a "probability distribution of probabilities," a concept of only apparent complexity, since the values of p can be thought of as simply the values assumed by a random variable. One deals, then, essentially with sequences of probability distributions.

It is now important to note that the events which determine the subject's "path" along the tree may be "experimenter-controlled" or "subject-controlled" or both. They are considered experimenter-controlled if the probability of applying one or the other operator does not depend on what the subject did on the last trial (the choice may be the experimenter's or determined by a chance device selected by the experimenter). They are considered subject-controlled, if what the subject did on the last trial completely determines the operator to be applied next. They are both experimenter and subject-controlled, if the choice of operator depends both on what the subject did and a chance device (or experimenter's decision). These distinctions are important not merely for classifying experimental procedures but for more fundamental theoretical reasons. One is interested in expressing several statistics of the p-value distributions, such as means, variances, higher moments, and others, in terms of the parameters of the models. Whether this can be done in terms of closed mathematical expressions or whether one must content oneself with recursion formulae often depends on whether the events are experimenter-controlled or subject-controlled. The same considerations are important in examining the opportunities for generalizing the model, say to more than two operators or more than two responses. It turns out that in general the situations with experimenter-controlled events allow the calculation of the moments of the distributions as closed expressions, whereas the others do not.

These considerations are important in that they reflect an attitude in research which views a task as requiring *integrated* efforts of the mathematician, the experimenter, and the statistician. To be effective, a mathematical model must be such that its conclusions and predictions are made in a language that can be easily interpreted. Cumbersome masses of symbols offered by the mathematician are useless unless some definite quantitative relations among variables can be discerned in them. The statistician should be able to say in advance what he can conclude from what kinds of data. The experimenter, most likely a psychologist, wants to learn something about the learning process. But if all of them are to work together in a coordinated effort, the needs of all must be considered. The evaluation of these needs and a familiarization with the mathematical research tool are the goals of the painstaking preliminary mathematical investigation. It is a reflection of the realization that the mathematician's work in theory is not merely the turning of a crank of a "deducing machine" but ideally should contribute to the mastery of the experimental situation itself.

There are other conditions, besides experimenter-controlled events, which introduce valuable simplifications into the mathematical apparatus. Each of these is subjected to analysis by the authors. As an example, consider the situation with two responses and two operators, so that the model is

$$\begin{aligned} Q_1 p &= \alpha_1 p + (1 - \alpha_1)\lambda_1 \\ Q_2 p &= \alpha_2 p + (1 - \alpha_2)\lambda_2 \end{aligned} \qquad [4]$$

Suppose now the learning experiment is so designed that the first outcome has the same effect on the first response as the second outcome has on the second response. An experiment in which such an assumption is reasonable involves guesses as to which of two light bulbs will go on. Guessing and failing to guess can be expected to have the same respective effects on subsequent guesses regardless of whether the right or the left

light bulb was involved. Translated into the conditions on the operators, this states that Q_1 should transform p in the same way as Q_2 transforms $q = 1 - p$, or, equivalently,

$$\alpha_2 = \alpha_1 ; \qquad 1 - \lambda_2 = \lambda_1 \qquad [5]$$

The first of Equations 5 is called the "equal α-condition." It turns out that when this condition obtains, explicit formulae for the moments (at least the first two) of the successive p-value distributions can be calculated even in the case of subject-controlled events, whereas only recursion formulae could be so calculated in the general case. Unfortunately, it is difficult to reconcile the equal α-condition with an experiment involving subject-controlled events (in the example cited, they were experimenter controlled), where the simplifications are of greatest service.

Mathematical simplifications occur also where one or both of the parameters λ are 0 or 1. In particular, assuming that $\lambda_1 = \lambda_2 = 1$ is equivalent to assuming that perfect learning can occur, that is, a complete fixation of one response and complete extinction of the other. This turns out to be quite a reasonable assumption in many experiments, at least over the ranges of trials considered. It is introduced in the experiment to be described below, with the result that the model can be "milked" much more extensively than in the general case.

Other simplifications result from a variety of other mathematical restraints on the parameters, for example the commutativity of operators, equality of the λ's, having one operator the "identity" operator (leaving p invariant), etc. It is the discussion of the mathematical consequences of these simplifications, their interpretations in terms of assumptions concerning the experimental situations, and the opportunities they provide for "milking" the model, that is, for drawing additional implications to be tested, that form the bulk of theoretical discussion in Part I.

Having investigated the mathematical apparatus at their disposal, the authors proceed to applications (Part II). Another full chapter, however, is devoted to a continuation of the theoretical discussion, namely the problem of relating the mathematical tools to the data.

A central problem, as one would expect, is the problem of identification, that is, the problem of deciding which concepts shall stand for which observables. Some such identifications have already been implicitly made. For example "Trial" has been identified with the situation where the subject has "an opportunity for choosing among a set of mutually exclusive and exhaustive responses." This gives a situational meaning to the index n in the stochastic equations. "Learning" has been identified with the changes in the probabilities of the responses and thus associated with the operators Q_i, etc. Other identifications are not so obvious. For example, in a two choice situation, the outcomes may be "right" and "left." But as modifiers of response probabilities, it is possible that "Same as last time" and "Opposite from last time" may be the more relevant events. The particular identification of events may or may not make a difference in the utility of the model. Whether among the several possible identifications, one, more than one, or none turn out to be successful, the methodological problem remains prominent: the search for identifications and the interpretation of the resulting successes or failures of the model.

Next the authors make a cursory attempt to relate their model to psychological theory. They take cognizance of two current theories of learning, the reinforcement theory, in which the outcomes have a direct positive or negative effect on subsequent probabilities of response (as "rewards" and "punishments"); and the contiguity theory, where the linkage of stimulus to response becomes established simply through repeated association. It turns out that the model offered by Bush and Mosteller is applicable to both psychological theories and so cannot in general mediate between them. Nevertheless, the choices of particular values of the parameters is in certain instances determined by whether one assumes the reinforcement or the contiguity point of view. It appears, therefore, that there are cases where a decision can be made on the basis of the model for the one or for the other extreme, or for

a compromise position where both factors operate.

Still another problem is that of estimating the parameters. In contrast to determinate models, where the estimation of parameters is a fairly routine procedure, the problem of estimation in a stochastic model raises questions of considerable theoretical importance.

To illustrate, consider the following learning curve, derived from extremely simple assumptions[2]

$$W = A(1 - e^{-kn}) \qquad [6]$$

where W is the number of errors, n the number of trials, A and k constants. Simple as the assumptions are, it should be possible to fit the resulting two-parameter curve to many sets of data. It is a trivial problem to "estimate" the parameters A and k: one simply reads off from the data the values of W for two values of n, which gives two equations in two unknowns to be solved for A and k.

Such a procedure cannot be applied to the Bush-Mosteller models, because their variables are not observables but real probabilities. Moreover they are probabilities not of measurable quantities but of other probabilities. Theirs is, in fact, a double estimation problem. The probabilities in question can be observed as frequencies only if one handles large aggregates of large aggregates, or, to state it backwards, one imagines taking samples of samples. Accordingly considerable enlightening discussion is devoted to this problem. Again, as in the case of the derivation of the relevant statistics from the model, the authors proceed by indicating how the estimation problems are simplified when the parameters can be assumed to have certain special values or to satisfy certain special conditions. Since these special cases of the general model are translatable into either particular experimental designs or into certain assumptions about the learning process, one sees once again the central theme of the book: an attempt to integrate the needs of the theoretician with those of the experimentalist, that is, to design a theoretical apparatus and an experimental situation such that each can be of the greatest possible use to the other.

[2] Let it be assumed, for example, that each error results in a constant decrement from an initial probability of error. In mathematical language this reads

$$\frac{dW}{dn} = A - kn$$

whose solution, assuming the initial condition $W(0) = 0$, is given by Equation 6. For several analogous examples, see, Gulliksen (1).

The applications

The actual application of the theory to specific experimental material begins with Chapter 10. The experiments examined include free-recall verbal learning (humans), avoidance training (dogs), imitation behavior (young humans), binary symmetric choices (rats and humans) and runway experiments (rats). In the last mentioned, the identification problem had to be raised *de novo*, and a theoretical model somewhat different from the one discussed is introduced. Any of the experiments should be sufficiently indicative of the method, and so the learning of avoidance experiments on dogs (Solomon-Wynne) will be discussed.

Essentially the experiment is interpreted as a realization of response probabilities in terms of frequencies. The mob of "identical" subjects needed for such realization consists of 30 "mongrel dogs of medium size," whose differences are supposedly compensated for by "handicapping" the task to be learned (the height of the barrier to be jumped is adjusted to the height of each dog).

The dogs are to learn to jump over a barrier to avoid an electric shock, warning being given by extinguishing lights. If a dog jumps between warning and shock, he is credited with a correct response, otherwise not. The effect of "success" on p is given by Q_1 in Equation 4 above; the effect of "failure" by Q_2. As is expected, the proportion of failures decreases fairly steadily, until, after the 24th trial no more failures occur even if the experiment is continued over hundreds of trials. One therefore assumes fixation of the jumping response both by the reward of avoidance and by punishment of shock and sets $\lambda_1 = \lambda_2 = 1$. The equations can now be written in a simpler form:

$$Q_1 q = \alpha_1 q$$
$$Q_2 q = \alpha_2 q \qquad [7]$$

where now q is the probability of the wrong response, α_1, is the "avoidance parameter," that is, the factor by which the probability of wrong response is reduced by each correct response, and α_2 is the "shock parameter," that is, the factor by which the probability of wrong response is reduced by the shock associated with it.

The task is now to estimate these parameters α_1 and α_2. This would be quite simple if the successive q's were known; but they too must be estimated. Nor can the q's be estimated simply from the successive numbers of dogs who fail to avoid the shock, since the sequences of responses for each dog may be different, and so the probabilities of avoidance for different dogs may be different at any particular trial. It is time to invoke the tricks of the trade of statistical estimation. The authors apply a battery of four methods, each a check on the others. They begin by estimating the shock parameter α_2. The first estimate is perfectly straightforward: since all thirty dogs failed to avoid shock on the zeroth (initial) trial, the fraction who failed on the first trial gives an estimate of α_2. Next, it is argued, that the fraction of those failing on the first trial who failed also on the second trial gives an estimate of $(\alpha_2)^2$. One could, of course go on in this way, obtaining estimates of the successive powers of α_2 and therefore α_2. But the authors take advantage of the machinery they had developed to attack the estimation problem from various angles. In Method 2, they use another statistic, namely the mean number of trials before the first avoidance, which, *under the special restraints on the parameters valid in this particular experiment*, allows them to use an alternative method of estimating α_2 and moreover a check on the estimate through a comparison of the observed standard deviation of the number of trials before the first avoidance with that calculated on the basis of the estimate! There are two other methods of estimating the shock parameter, which it suffices to mention. At any rate, the purpose of the high powered analysis of the model now becomes clear. It establishes an apparatus for "milking" the model, i.e., for deriving a great many implications about the statistical aspects of the data and thus for obtaining many checks, double checks, cross-checks, etc. for the "validity" of the model. The use of "stat-dogs," that is, of a chance device which behaves according to the prescripts of the stochastic model, allows the comparison of even more features of the data with those "derived" from the model, for example the means and the variances of the number of trials before the second avoidance, total shocks, trial of last shock, alternations, longest run of shocks, trials before the first run of four avoidances, etc., etc. All these statistics would have been mere curiosities without the model. Here they are put to work as a battery of tests. Agreement is excellent throughout.

A learning curve (cumulative number of shocks against trials) is also derived and fits the data as snugly as can be desired. A determinate model would have to be content with just this agreement. Thus the greater richness of implication of the stochastic model as compared with the determinate model seems demonstrated.

The authors do not draw this conclusion, since they do not discuss the determinate model at all. They do indicate in their concluding remarks the vast difference between their theoretical approach and the purely empirical statistical evaluation of a set of data:

> Suppose a learning experiment, involving 1000 subjects for 300 trials each, has recorded either a success or a failure for each subject on each trial. These data are turned over to a clerk for summary analysis. There are now literally thousands of independent questions about these data. . . . Examples are:
>
> 1. What is the over-all percentage of successes?
> 2. What are the trial-by-trial success percentages?
> 3. What is the best-fitting cubic curve that can be drawn through the trial by trial percentages?
> 4. What is the mean trial number on which the last failure occurred (how about the median instead of the mean)?
> 5. What is the mean trial number on which the first success occurred (the second, the third . . .)?
>
> . . . . . .
>
> Even this small group of questions makes clear the rich variety of possible summary questions

> that can be asked of this simple data sheet. A complete summary could clearly take hundreds of pages. The clerk working over these data would just have to compute a new number in response to each question. . . . Knowing answers to one hundred questions would not help appreciably in answering the next hundred.

The point is driven home by the appropriateness of the word "clerk." The implication is that the answers to the questions can be given without any insight into their "meaning" and that the model, by indicating how all such answers become consequences of a few simple assumptions, do provide the "meaning." The authors invoke a comparison with Kepler's laws, from which an enormous wealth of astronomical observations could be deduced as consequences. (By not invoking a comparison with Newton's laws, of which *Kepler's* laws are consequences, the authors reveal both their modesty and their hope of further theoretical generalizations to come.)

And so the greater power of the model compared with mere quantitative description is amply demonstrated. It seems more interesting, however, to make a comparison between the relative merits of different *theoretical* approaches. We have already pointed out the advantages of the stochastic model over the determinate one, in that only the former is able to predict a great many statistical relations in a particular set of data. However, this does not settle the question of the relative utility of the two types of model. There is more to the determinate model than meets the eye.

We must confine ourselves to the barest outlines of an argument. Recall the simple-minded determinate model described above (p. 65). Since the derived learning curve (Equation 6) involves two free parameters, it should not be difficult to fit the curve to several sets of data by adjusting the parameters. However, the parameters A and k are not merely numbers to be adjusted: they have an interpretation. The parameter A measures the initial fraction of errors per trial; k, essentially the effectiveness of "punishment" in eliminating errors. The first parameter can be considered to be under the control of the experimenter, since the initial fraction of errors can be established by the experimental situation. (Thus one can make any desired fraction r/s the initial frequency of errors by presenting the subject with s equiprobable choices, $s - r$ of which are correct). Having determined the values of A and k by fitting one learning curve, the experimenter can *manipulate* A and thus put the model to a further test. In principle, he can also manipulate k by changing the "punishment."

The more free parameters there are in a determinate model, the easier it is to fit numerous sets of data with appropriate curves, but the less significant theoretically is the "success" of the model. However, if the parameters are *manipulable*, theoretical significance can be restored. In principle, some of the parameters of the stochastic models described by Bush and Mosteller are also manipulable (i.e., the shock parameter in the experiment described). However, the center of attention is directed away from the manipulation of parameters and toward a detailed analysis of the stochastic situation at hand. Herein lies the meaning of the remark made above that the stochastic model is characterized by depth rather than by breadth.

Similar remarks apply to the theoretical interpretations of the stochastic model. The authors make a few departures into what they call "pseudo-neurology." That is, they give their equations a "physical" interpretation in terms of interactions of conditioned and unconditioned "elements" of the nervous system during the learning process. This departure, however, is in no way germane to the central argument. It is somewhat analogous to the attempts by early investigators of electromagnetic fields to represent the fields by mechanical models. There is some justification in refusing to speculate about neurological bases for behavior, but also some justification for pursuing such speculations. Granted that it is all but hopeless at this time to trace or even to infer what actually happens in the brain during learning, relatively simple-minded "pseudo-neurological" models can suggest a wealth of hypotheses to be tested.

Such has been the experience of the workers associated with Rashevsky and Landahl at Chicago.

As an example, we can examine Landahl's theory of learning. It contains three basic ideas: a postulated law of transynaptic transmission of "excitation" and "inhibition" (from Rashevsky's earlier investigations); a circuit of hypothetical neural connections; and a postulate concerning random fluctuations of thresholds. It is important to note that all three postulates had served in the construction of models involving *other* phenomena, e.g., reaction times, discrimination of intensities, etc. From this "physical" (or, as Bush and Mosteller would prefer to call it, "pseudo-neurological") model, Landahl derives a learning curve involving several free parameters (3). However, because of the "physical" basis for the model, the parameters too have a "physical" interpretation. To the extent that the parameters are associated with individuals or groups of individuals and to the extent that they can remain undisturbed by the learning situation, they allow the prediction of the behavior of the same individuals in other situations. In another learning model, Landahl has actually followed this procedure and was able to predict a complete learning curve after all of its free parameters were "eliminated" by changing the experimental situation (2).

It must be stressed that *in principle* the introduction of additional parameters and their experimental manipulation is applicable to the stochastic model as well as to the determinate one. Yet the emphasis of the stochastic model is such that this procedure seems less feasible, than in determinate models. The latter deliberately sacrifice the refined analysis of the stochastic process (confining the attention to crude time averages) in order to gain flexibility by handling a wide variety of assumptions. The stochastic model of Bush and Mosteller can also be viewed as a result of sacrificing some aspects of the learning situation to gain greater insight into others. In particular, individual differences are deliberately neglected (or an attempt is made to eliminate them). Any such differences in the objective situation must appear in the stochastic model in the "unaccounted variances." In the determinate model, on the contrary, individual differences are reflected in the parameters, while the chance events of learning appear as "unaccounted discrepancies." One might argue that where learning takes place over hundreds of trials (as, for example, in experiments reported by Gulliksen [1]), determinate models are probably appropriate; where only a few trials are involved (as in the Solomon-Wynne experiment), the stochastic model is more useful. At any rate, it is fortunate to have the means to attack the same problem by roughly complementary methods. The far-reaching analysis of the stochastic model by Bush and Mosteller is a welcome contribution to the growing stock of mathematical tools available to the behavioral scientist.

REFERENCES

1. Gulliksen, H. A rational equation of the learning curve based on Thorndike's law of effect. *J. gen. Psychol.* 1934, 11, 395–434.
2. Householder, A. S. & Landahl, H. D. *Mathematical biophysics of the central nervous system.* Bloomington, Ind.: Principia Press, 1945.
3. Landahl, H. D. Studies in the mathematical biophysics of discrimination and conditioning. *Bull. math. Biophysics*, 1941, 3, 13–26.

ROUND TABLE DISCUSSIONS ON BEHAVIOR THEORY

Physicians often publish in their journals verbatim transcriptions of clinical pathological conferences in which they make public the process by which they come to their diagnostic conclusions and carry out other clinical functions. Lawyers often put complete transcriptions of their legal reasoning on the record. With scientists this is not so common. In this department we expect to present edited segments from round table discussions concerning behavior theory. These discussions will be largely informal and unrehearsed, and they will illustrate more the activity of the development of such theory and the interplay among individuals and disciplines, rather than finished products. At points their value will more likely reside in the liveliness of the interchanges rather than the accuracy of the conclusions.

The Cost of Decision Making

Formal Models Seminar, March 30, 1954 and April 13, 1954[1]

Participants (at time of the discussion, all associated with the University of Chicago except as indicated): Jay B. Best, physiology; Robert I. Crane, history; Donald W. Fiske, psychology; George R. Flowers, biology; H. G. Landau, mathematical biology; Jacob Marschak, economics; Robert S. Merrill, anthropology; James G. Miller, psychiatry and psychology; Andreas Papandraeou (University of Minnesota), economics; Anatol Rapoport, mathematical biology; Marianne Simmel (University of Illinois), psychology; and Burton Stein, history.

Marschak: What is the value of the information we may obtain, say, from a spy or from a scientist? What relations can be found between value and amount of information?

Take the case of making even-money bets that a blue ball will be extracted from an urn containing red and blue balls. We may simply use our knowledge of the probability that the ball will be blue, or we may supplement this knowledge by using a spy. Let p = the probability of getting a blue ball; then $1 - p$ = the probability of a red ball. Set the amount of one dollar to be gained or lost per decision. Let

$$p > \tfrac{1}{2}.$$

[1] This discussion was held at the University of Chicago by members and guests of the Committee on Behavioral Sciences. The discussion on March 30 was initiated by Jacob Marschak and on April 13 by Jay B. Best.

Without the informer, we shall always bet on blue; with the informer, we shall always gain a dollar. Compare the two average gains:

The value, $U_0 = (+1)p + (-1)(1 - p) = 2p - 1$

(without the informer)

The value, $U_1 = 1$

(employing the spy)

The *value of information*, then, is

$$V = U_1 - U_0 = 1 - (2p - 1) = 2(1 - p)$$

Figure 1 is a plot of the situation when the spy is reliable.

Suppose the spy is unreliable. Let q = the probability of his telling the truth. Then

$$U_0 = 2p - 1$$

$$U_1 = (+1)q + (-1)(1 - q) = 2q - 1$$

Therefore, $V = 2(q - p)$ of which the previous formula for V was a special case with $q = 1$.

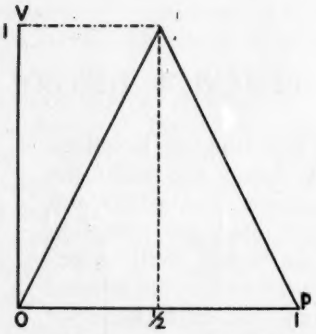


FIG. 1. Value of Information when Spy Is Reliable

Further, the net value gained by using the spy is

$$U_1 - U_0 - c$$

where c is the cost paid to the spy.

Some spies, or other sources of information, are more costly than others. Some are more reliable than others. We are thus faced with choices among various net values:

No spies: $2p - 1 - c_0$

Spy No. 1: $2q_1 - 1 - c_1$

Spy No. 2: $2q_2 - 1 - c_2$

Note that if the spy is paid only when he's right, then the net gain is $2q_1 - 1 - c_1q_1$, etc. (A suggestion of A. Rapoport.)

We have seen the *value of information* from a completely reliable spy is $2(1 - p)$. This is different from the *amount of information* (Shannon). The amount of information doesn't depend on the economic value to the decision maker. It depends entirely on the properties of the probability distribution and is defined as

$p \log p + (1 - p) \log (1 - p)$,

or the antilog: $p^p(1 - p)^{1-p}$

Shannon's formulation plots into:

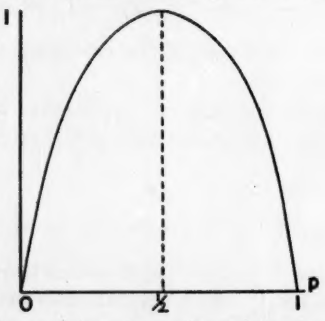


FIG. 2. Shannon's Information Plotted against p

Thus, in our case, value and amount of information coincide for the case of $p = 0$, 1, and ½, but not for the intermediate cases.

A farmer is faced with the problem of how many acres to cultivate. The concept of action will be introduced here. At this point, action and decision will be considered as identical.

Let a = acreage, or action
μ = mean price of grain
x = deviation from μ

Then price = $\mu + x$

and $Ex = 0$ (that is, the expected deviation is zero). Write $Ex^2 = \sigma^2$ (= variance).

Let $u(a, x)$ = profit.

Let us say that in Condition 0 the farmer has no information, i.e., he doesn't know the price; in Condition 1 the farmer does have a source of information. (A good farmer actually "employs two statisticians": one for the market, one for the weather.) Let us assume yield/acre is always = 1. We shall compare the highest expected profit (highest long-run average profit) that can be achieved in either of these two conditions, and compute the value of information:

$$V = \mathrm{E} \max_a u(a, x) - \max_a \mathrm{E}u(a, x).$$

The revenue will be $a(\mu + x)$. Suppose he farms different acres at a different expense. As a simple assumption, let his

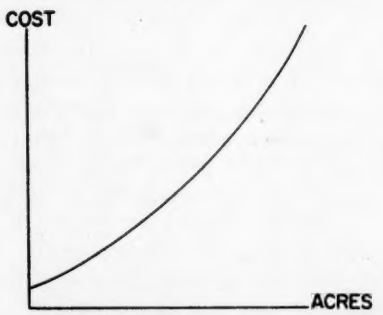


FIG. 3. Cost increment with increased number of acres

total cost = $(\frac{1}{2}a^2 + \beta a + \gamma)$. The derivative of this is $a + \beta$; that is, the additional cost of adding acres is a linear function of the number already under cultivation. The profit, u = revenue less cost.

Now consider Condition 0, wherein the farmer has no information:

Eu (or long-run average profit)

$$= Ea(\mu + x) - (\tfrac{1}{2}a^2 + \beta a + \gamma);$$

$$Eu = -\tfrac{1}{2}a^2 + a(\mu - \beta) - \gamma \quad (\text{since } Ex = 0) \tag{1}$$

The derivative of this is: $-a + \mu - \beta = 0$, if Eu is at its maximum with respect to a.

$$U_0 = \max_a Eu = \frac{(\mu - \beta)^2}{2} - \gamma,$$

which is the maximum average profit that can be obtained when x is unknown.

Now let's consider Condition 1, wherein the farmer has information on prices. First, find

$$\max_a u(a, x)$$

He tries to maximize u, given x, by choosing the best decision a_1. We have $u = -\frac{1}{2}a^2 + a(\mu + x - \beta) - \gamma$. This is similar to the right-hand side of Equation (1) except that $(\mu + x)$ is substituted for μ. In this second case, differentiation gives us:

$$-a + \mu + x - \beta = 0,$$

when u is at its maximum with respect to a. Then

$$\max_a u(a, x) = \frac{(\mu + x - \beta)^2}{2} - \gamma.$$

Now we want to find the expected value of this maximum profit

$$U_1 = E\frac{(\mu + x - \beta)^2}{2} - \gamma$$

$$= \frac{(\mu - \beta)^2}{2} - \gamma + \frac{\sigma^2}{2}.$$

(The last term is one-half the variance of price.)

What is the value of information?

$$V = U_1 - U_0 = \frac{\sigma^2}{2}$$

Rapoport: The result would be some other multiple of σ^2 if Marschak had not chosen $\frac{1}{2}$ for his original equation.

Marschak: Yes. In any case, the value of information is related to the variance (expected square deviation) if profit is quadratic in the "decision variable" a. This result is independent of the distribution of the "nature variable" x. But notice that when this distribution is normal, Shannon's measure of the *amount* (not the *value*) of information about x is also related to the variance of x.

Landau: The acreage is actually finite; $a \leq A$, say.

Marschak: In this case, the optimal a is the same as in my case, or is equal to A, whichever is smaller.

Miller: What is the relationship of what you've said to organization theory?

Marschak: I have taken a simple case with one actor and one observer (several observers—spies—in the case of the urn). In the general case, the problem is whether or not we should maintain a given line of communication, that is, when does its cost exceed its value? Later we'll get into cases with multiplicity of actors. You may be interested in today's problem in its relation to the evolution of the brain. The value of information may be nil to a species that lives in an environment with very low "variance," e.g. receives its food always floating in the same direction. Such a species need not have sensory organs.

• • •

Best: My comments here stem from the discussion by Marschak in the formal models seminar of March 30, 1954. A certain price per decision was exacted by a "seer" or observer with regard to the number of acres a farmer should plant in order to "win" in the face of changing market conditions, etc. It was assumed that the observer was not always correct but merely had a higher probability of being correct than the farmer trusting to chance would have. Viewing animals as an economic structure in which the brain and sensory organs would be analogous to the seer and the metabolic requirements for their maintenance the cost, Marschak raised the question of whether animals that were susceptible to impinging stimuli from all sides, such as those that live in the sea or calm water, did not have larger brains than those that live

in a stream where all significant stimuli come from upstream.

As a rule which seems to be fairly valid, those organisms which obtain their food by foraging, i.e., moving about and seeking out other organisms to prey on and eat, have developed the largest brains and head ganglia with the most highly specialized and differentiated receptor systems. An example of this is that the Cephalopods in the phylum of Molluscs, by an entirely independent evolutionary route from that of the vertebrates, have chosen a predatory and highly motile mode of existence with concomitant development of the head ganglion into a fairly respectable brain and eyes that closely resemble those of vertebrates. This development stands in contrast to that of the other molluscs who lead more sessile existences and have much less extensively developed receptor, conductile and integrative tissues.

The biological question posed by Marschak might be approached by rephrasing the problem slightly from the form chosen in the seminar discussion. An animal, by virtue of the fact that it goes abroad in search of food, is confronted with a larger domain of possible events, both profitable (nutritious) and dangerous, than if it followed a more sessile pattern of behavior. One might suppose that an organism is faced with n_0 decision situations *per unit time* if it sits still, and $n_0 + M$ decisions per unit time if it is moving about, where M would be some monotone function of its speed and range of movement. Suppose further that

p_0 is the probability of a correct decision by "chance."

$1 - p_0$ is the probability of a wrong decision by "chance".

w is the amount the animal "wins" on a correct decision.

L is the amount the animal "loses" on a wrong decision.

If the animal has a brain we might suppose it to make correct decisions with a probability of p_1 and wrong decisions with a probability of $(1 - p_1)$ and that $p_1 > p_0$. The metabolic cost of maintenance *per unit time* of such a brain will, however, be g, and this will, in general, take more the form of a "fixed overhead" cost rather than an assessment per decision for the animal. Then for a motile animal with no brain the net nutritional gain per unit time would be

$$U_0 = wp_0(n_0 + M) - L(1 - p_0)(n_0 + M)$$

with the sessile existence being represented by the special case in which $M = 0$. For a motile animal with a brain, however, the net nutritional gain per unit time would be

$$U_1 = wp_1(n_0 + M) - L(1 - p_1)(n_0 + M) - g.$$

The nutritional gain by virtue of having a brain is then just

$$U_1 - U_0 = (w + L)(p_1 - p_0)(n_0 + M) - g,$$

so that possession of a brain will represent a handicap unless

$$(w + L)(p_1 - p_0)(n_0 + M) \geq g.$$

One may look upon the ratio

$$E = g/(p_1 - p_0)$$

as a kind of "extravagance index" associated with the possession of a brain. Then a brain will represent a handicap unless

$$(w + L)(n_0 + M) \geq E.$$

Thus, one has the common-sense result that the higher the stakes $(w + L)$ and the more the animal "gets around," the more it can afford to pay metabolically for being intelligent.

Miller: Instead of defining p as the probability of the correctness of decision, we might define it as the probability of an adequate defense made by an organism in resistance to environmental stresses. This makes it a somewhat more general formulation. For a subsystem dealing with stress to be worthwhile, $U_1 - U_0$ must be greater than zero. If not, the subsystem is extravagant. Examples would be an FBI turning into a Gestapo, automatic headlight-dimming devices in automobiles which discharge batteries, etc.

Marschak: A better title for Best's discussion would be "Gain and Cost Due to Decisions and Observations." But this is

not very important. It seems to me very striking that this group readily accepts Best's notion of *molluscus economicus* while it resists that of *homo economicus*! There were objections when I spoke of humans as entrepreneurs maximizing something (such as dollars), but there have been no objections today to dealing with molluscs in this way! I was interested to learn from Best's talk that biologists do think in terms of what may be called the economics of evolution. Best, you talk as if animals were trying to economize on energy?

Best: Yes. But I like to view this process from the standpoint of the species rather than of the single animal. For a species which has a high rate of division and growth, it actually may be more economical to take a loss in individual lives so that the species may evolve. But this is by no means a unique way of doing business. Humans, for instance, have an elaborate nervous system for protection of *individuals* in the species. Still other survival processes are also successful. I think species maximize probability of survival of their genetic descendants through adopting a mode of life best suited to this end. Suppose an animal has adopted a particular mode of life: Is it then useful or not for it to have a brain? Best strategy for the mollusc involves getting as much food as is necessary. Those who do this best will survive.

Marschak: In the simple case of either surviving or not surviving, you may say the utility function is two-valued. In this case, to maximize the probability of survival is formally identical with maximizing the expected utility, where the utility is two-valued: utility of survival = 1, utility of nonsurvival = 0, expected utility = $p \cdot 1 + (1 - p) \cdot 0 = p$.

Best: I wasn't talking about a decision. Suppose you start off with a group of molluscs from which equal numbers of different evolutionary strains will be drawn. Those which enjoy the greatest net caloric intake will survive.

Marschak: Would you say that you can interpret evolution by saying that species maximize the rate of population increase? The balance of births over deaths should be, percentagewise, as high as possible. It does not follow that the more complex organisms will be those that survive. Have they, in fact, increased their share in total living matter?

Best: I think it is Dr. Gerard who is interested in the idea of survival of more complicated forms. I think one can point to evolutionary development towards less complicated forms, e.g., the evolution of parasitic forms from animals which were once free-swimming.

Miller: I think Dr. Gerard has modified his thinking about this notion during the past year. He would probably say, though, that probably the more complicated organisms came after the less complicated organisms. It doesn't mean necessarily that less complicated forms will disappear; if environmental stresses are minimal, they will survive. In general, however, when mutations occur, you can state that there are three possibilities: (*a*) a less complicated structure mutates to a more complex one, (*b*) a structure mutates to one of equal complexity (or efficiency), (*c*) a structure mutates to one of lesser complexity (or efficiency). Actually, complexity and efficiency don't necessarily correspond, so perhaps there are six rather than three possibilities: more complexity with either more or less efficiency, and so on. Increasing complexity in evolution must be considered in relation to the nature of the environment, to determine whether it will lead to survival. A human being is capable of surviving in a more complex environment than an amoeba.

Marschak: Probably you include our social institutions, e.g., our municipal transportation, among our "organs." But, then, are we not more helpless than animals in a changing environment, as when elevators and water heaters stop functioning in a war or a strike?

Miller: Most discussions I've read of social evolution have been nonsensical because they deal with vague analogies rather than with precise study of the mechanisms. I think it is possible, however, to talk sensibly about social evolution. If an individual can't handle stresses, and has become a specialized part of a group, he is now more vulnerable. But he has gained many advantages. A good analogy to the

Marschak-Best model is the assignment of sentry functions in a wagon train passing through Indian territory; it is a more effective organization than any individual can be.

Simmel: Simpson uses an interesting analogy on evolution. Rocks, stones, and grains of sand are placed in a bin. The large rocks are pace-setters; after they are put into the bin, there are still spaces for the smaller stones, and thereafter for the sand. If there are changes in the properties of the pace-setters, then perhaps "mutants" in other categories have better chances of "survival."

Marschak: Thus, species form an environment for each other.

Best: An example is that a species dependent upon photosynthetic plants for food can't multiply out of proportion to a certain rate of photosynthesis.

Merrill: To go back to Best's model: No account was taken of the overhead on the locomotor system (as well as the brain). Could this model be modified to include the matter of change, say, between the locomotor system and the brain? In anthropology, the current view is that locomotor change (e.g. man's emergence as a bipedal animal) is prior to the brain change.

Best: Actually, locomotor system is involved in my term M.

Crane: In the illustration of the pioneer group in Indian territory, one must weigh the advantages of such an organization against the advantages of the passage of a single scout through this territory: he can remain unseen by the enemy.

Miller: This is analogous to defenses of a biological organism: a small, weak organism can be camouflaged—not so a larger, stronger organism.

Crane: I took it that Best viewed this evolutionary process as one of maximizing the probability of survival of the larger number of offspring.

Best: Usually by successful survival one means that the species has existed for a long time.

Crane: Some species are so successful in maintaining their existence and in breeding that eventually their numbers increase to a point where sufficient food is unavailable. Consider the lemmings. Such success is illusory.

Best: If there are so many lemmings that the only place to put them is in the ocean, then lemmings are successful. I would view as successful up to the present all lemmings that you find. (In answer to Marschak) No, I wouldn't consider success in terms of total mass or number or health of organisms.

Fiske: Rather the number of generations the species has kept going.

Best: Yes. Probability of survival of their lineal descendants.

Fiske: Using this criterion, what you are maximizing is not rate of growth of the population but the chances of the rate staying positive.

Best: That's right. We got off into a subcorollary of this: Does increase of the total number tend to increase the probability of continued survival over a long period of time?

Marschak: You're interested in the integral of population over time: the total number of molluscs that have lived until now. You have tried to maximize this.

Best: It would be a question whether you would want to take the integral over all those that have existed or wished to count the total time over which they have continued to exist.

Flowers: I'm wondering about what you mean by correct decision by which the animal "wins" the amount W. Do you consider that he has a net gain of energy after his correct decision or that he made a correct decision in that he got his prey but may or may not have a net gain of energy?

Best: If the prey is too small, perhaps there's no gain.

Flowers: He may intercept the prey but nutrition may be insufficient.

Marschak: What is Best's payoff matrix? Is it like this? In the case of motile animal we have (denoting, as before by a the decision, or active, variable, and by x the nature variable, or environment):

	$x = 1$ = Danger	$x = 2$ = Food
$a = 1$ = run	$W - g$	$L - g$
$a = 2$ = grab	$L - g$	$W - g$

Now consider the sessile organism:

	$x = 1$ = Danger	$x = 2$ = Food
$a = 0$ = sit	L	W

Flowers' objection is that the situation is more differentiated than the two symbols W and L permit us to describe. There are more than only two possible outcomes.

Flowers: Yes. Moreover, the organism doesn't have information to know whether he's going to spend more energy than he's going to take in.

Marschak: Would you suggest that there are more than two actions in the first case and more than one in the second case? You would complicate this further?

Best: This is not a pretentious model but merely a reformulation of the mathematical statements made by Marschak.

Simmel: One difficulty with the model is that it defines payoff of a single individual decision. But what matters is the payoff over a long time for the whole species. If the organism grabs when there's danger, its fate is settled once and for all, while if the organism runs toward food or away from danger on a single decision, its fate is not settled once and for all.

Marschak: Yes. This matrix may be considered set up for one unit of time. A more complete model would assign a payoff to each sequence of outcomes. In business, you compute the sum of discounted values of future profits, and maximize this. However, Best will form a sum of payoffs, not only for each individual over time, but also compute the sum of such sums over all individuals of the species.

Miller: Coming back to the discussion between Crane and Best, I'd like to raise a *caveat* about a word like success. Success is a value term and as such must be carefully inspected. Perhaps to the biologist, continuation of the species is the most successful accomplishment; however, to others, continuation of the individual is more important. It may well be that one is incompatible with the other. It may also be that a lemming would not consider himself successful in Best's terms. Propagation might go on so rapidly in India, say, that the next generation would not have the same opportunity to propagate as the last generation simply because there wouldn't be enough food to sustain them. This suggests another conception of success. We may have at least five or six different values implied in the word success.

Marschak: Best's concern was only with prediction. He is not concerned with the mortality rate of Indian babies or of fish eggs, offset by large birthrate. He will say the Indians are all right as long as there are 400 million of them and as long as this number is stable. In fact, he is only interested in this number. He is able to predict that fish or people with a particular kind of device (high birth- *and* deathrate) maintain themselves. Perhaps it is sound for Best to stick to his particular criterion of survival probability as a tool of prediction even though Crane might rightfully object to it as a moral value.

Landau: The criterion of survival explains the direction of evolution, but the fact we seem worried about is that there are a lot of different ways of surviving. I would like to suggest that there may be a difference between organic evolution and social evolution. An organism doesn't really have a choice: If a mutation occurs, it may or may not be successful and consequently may or may not survive. And there may be several possible ways to survive; sessile and mobile molluscs both have survived. When Miller spoke of the division of labor among people crossing the prairie, it occurred to me that the humans would be more successful perhaps if each individual could perform all these functions (sentry, fighting, working) by himself. However, this would require mutation, and we have no control over this. The important difference, then, it seems to me, is that in social evolution we have some choice, or so it appears.

Marschak: To this extent, social institutions are artifacts, like machines. In order to achieve a purpose, you choose the best machine, i.e., one that maximizes something, e.g., profits. At the turn of the 18th–19th century intellectuals were busy discussing various political constitutions among which they had choice: see the Federalist papers.

Stein: I was wondering after Flowers'

question if the item food might not be considered net caloric gain. Perhaps this is closer to what Best was thinking of. If so, perhaps one of the prime functions of the added brain is to make the fairly subtle decision about whether there will be a net gain.

Best: Yes. Net caloric gain is closer to what I meant. But this makes the model less simple.

Marschak: Can you supply numbers for the matrix? Nobody will criticize you if they're not absolutely right. How about:

	Danger	Food
Run	$-1-g$	$-1-g$
Grab	$-100-g$	$+50-g$

If you run away from enemy or food, all that happens is that you've expended a little energy, say one unit; and g, in Best's terms, is the energy cost of maintaining the brain.

Miller: Perhaps Danger-Grab should be minus infinity. Though perhaps not if you consider the species as a whole.

Landau: It would only make sense in terms of evolution of species as a whole.

Marschak: You have to cumulate over time and over all individuals.

Landau: Anyone who could fill values into this matrix at all realistically would be making a major discovery.

Papandraeou: A question concerning the predictive value of this model. You indicated that prediction was the primary objective of it. In economics, where we also speak of maximization, our primary objective is to predict the signs of rates of change of variables with respect to parameters. For instance, if there is an increase in the wage rate by the union, the fact that the firm maximizes profits permits us to predict, given some other assumptions, that somebody will be fired. Now, what kind of prediction concerning pure behavior of the molluscs can you make from a model of this sort?

Best: One common-sense result mentioned was that the more the animal gets around and forages for prey, the more valuable a brain will be. At least among the molluscs, this happens to be so. The more active, predatory ones have so developed. It's admittedly a hindsight prediction.

Marschak: As in Papandraeou's wage example: If parameter g (brain's "wages") increases, brain becomes a waste. Another parameter is p_0, a property of the environment. If p_0 is ½, then it pays very much to have an added sensory organ. When it is near 1 or 0, then it will not pay.

Best: I set up my probability slightly differently than yours. This is the probability of obtaining food if it grabs at random. In your case, didn't you have an over-all climatological probability?

Miller: In answer to Mr. Papandraeou's question about what empirical studies can follow from this model: Another study would be the inheritance of acquired information —Lamarckianism vs. Darwinianism. One might study different generations of molluscs to determine the respective probabilities with which each generation grabs and runs. Perhaps the present generation would grab vs. run with the same probability as a previous generation; or maybe they would grab vs. run with a chance probability at first and gradually change as time went by. (McDougall did experiments on this problem a decade or so ago.) The question, then, is: Do you have a "spy" which gives you the information on the basis of which you decide to grab vs. run? Or is this information gained through inheritance or observation of parents' responses?

Crane: If it costs a sizable amount for a mollusc to maintain a brain and since there are the two types of molluscs—motile and sessile—why did the motile ones become motile and develop brains while the sessile oysters, for instance, got along all right?

Best: You have mutations occurring all the time. If you want to know how many ways there are of surviving, you can count the number of species there are on earth.

Landau: Simpson's rocks and pebbles analogy is pertinent here. Different mutant forms survive in different environments.

Miller: One may use a very simple figure of speech. If you have enough income so

that you don't have to work, and then you added another subsystem known as your mother-in-law, you have several possible courses of action: You can starve to death, or you can go to work, or you can send your mother-in-law out to work for you. So a mutation presents new species that evolve with new problems of survival and perhaps new assets for survival.

Marschak: I'd like to continue my answer to Mr. Papandraeou, by interpreting Best's results in terms of our previous two sessions. We ask how the variations in parameters affect the choice of the organization. We have to distinguish here two groups of parameters. One group describes nature, in terms of probabilities of its possible states. In Best's model, this was represented by p_0. Another group consisted of parameters characterizing the payoff, as a function of the decision (action) and the state of nature. To review:

Payoff function: $u(a, x)$, as represented by the matrix.

Here a = decision (it has, in our example, two values: run and grab);

x = state of nature (also two values: damage and food).

If the mollusc has a brain, its rule of action, $a = \alpha(x)$ is as follows: "if x = damage, than a = run; if x = food, then a = grab."

Best also mentioned that from the point of view of the species, it is the long-run average that matters. In his formulas, he multiplied each caloric payoff by its probability, and added. Clearly this long-run average (or "expected payoff") depends on: (*a*) these probabilities; (*b*) the payoff function; (*c*) the rule of action.

In the case of the mollusc with brain, the action a depends on information about x; but in the case, "without brain," the action (viz., sitting still) does not depend on x. It is, in fact, constant. We have thus two "information structures" each having one row and one column:

Without brain:

	x	(Action doesn't depend on the state of nature.)
a	0	

With brain:

	x	(Action does depend on the state of nature.)
a	1	

Here 0 stands for "independence of a from x," 1 stands for dependence. In my example of a farm given last time, there might have been two action variables: a_1 = decision about amount of grain and a_2 = decision about amount of livestock; and three nature variables were x_1 = the price of grain, x_2 = the price of livestock, and x_3 = the weather. A hypothetical information structure might be: to base planting on the knowledge of grain prices and of weather; and to base livestock raising on the knowledge of livestock prices:

	x_1	x_2	x_3
a_1	1	0	1
a_2	0	1	0

With this information structure the rule of action will be of the following kind:

$$a_1 = \alpha_1(x_1, x_3); \qquad a_2 = \alpha_2(x_2).$$

Among the various possible functions α_1, α_2 there will be some that are best. When they are used, the expected payoff less the cost of information equals the value of the above information structure.

The case of Best has an additional complication. He assumes an "imperfect spy"; his p_1 is less than 1. The case of the imperfect spy was discussed in my first session. In general, decisions aren't based on the true state of nature but rather what the decider thinks the state of nature to be.

Let y = information available to the decision maker, and

x = true state of nature.

Then $y = \eta(x)$, where the function η may be called the coding function. It depends on the organs or methods available to learn about the state of nature. The personnel manager of a firm observes the labor market, but he does not communicate all his observations in detail tc, say, the president. He'll make use of averages to give a more compact statement, which is, presumably, cheaper to communicate and easier to act upon. Thus, the rule of action is, in general, not $a = \alpha(x)$ but, rather, $a = \alpha(y)$, where $y = \eta(x)$.

In the case of an imperfect spy, $\eta(x)$ is a random function, viz.:

$$\eta(x) = x \text{ with probability } p_1$$
$$\text{non-}x \text{ with probability } 1 - p_1,$$

where p_1 has the same meaning as in Best's model. There may be a choice of the best, or cheapest, coding function, given the other conditions. It may not be worthwhile to get all information exactly. We ask: What kind of information is it important to require observers to communicate? Sometimes it is better to know a shorthand summary. For example, sometimes the decisions of the high levels of government depend on national income, the general price level, etc. All these are averages ("index-numbers") or totals, not details. Thus $\eta(x)$ is not equal to x. This is also the case when, for reasons of economy, one decides to accept a certain amount of "noise," by hiring an "unreliable" spy.

To take account of this generalization, we must describe the information structure of a given network by giving, for each decision variable (say, a_i), not—as I have done earlier—the list of nature variables that are exactly known to the ith decision maker; but, rather, by giving, more generally, the coding function $\eta_i(x)$ that indicates how the information available to the ith decision maker is formed from nature variables. The information structure is the n-tuplet $\eta_1(x), \cdots, \eta_n(x)$, if there are n decision variables. Clearly the previous definition of information structure is a special case: to pick out some of the nature variables and leave the others out as irrelevant to a particular decision, is a special case of coding.

In any case, our problem is to compare between two or more competing networks, each characterized by a cost and a certain information structure and, hence, characterized (given the payoff function and the nature's probabilities) by certain best rules of action. On this basis the net values of networks are compared. This was discussed in the previous session and has been also done, in essence, in Best's formulas.

(For a further exposition of ideas similar to those presented by Marschak in this discussion, see the following papers: "Towards an economic theory of information and organization," in *Decision processes*, edited by Coombs, Davis, and Thrall, New York: Wiley, 1954; "Elements for a theory of teams," *Management Science*, Vol. 1, No. 2, 1955, 127–137; "The firm as a team," [abstract] *Econometrica*, Vol. 22, No. 4, October, 1954.)

Writing, when properly managed (as you may be sure I think mine is) is but a different name for conversation. As no one, who knows what he is about in good company, would venture to talk all;—so no author, who understands the just boundaries of decorum and good-breeding, would presume to think all: The truest respect which you can pay to the reader's understanding, is to halve this matter amicably, and leave him something to imagine, in his turn, as well as yourself.

—Laurence Sterne, **The Life and Opinions of Tristram Shandy, Gentleman**

ABSTRACTS OF CURRENT LITERATURE WITH INTERDISCIPLINARY IMPLICATIONS

Prepared by Members of the Staff of the Mental Health Research Institute

1 *Baitsell, G. A.* **The cell as a structural unit.** *Amer. Scientist*, 1955, 43, 133–141.

A review of the studies of the last forty years on molecular pattern determination is presented, from the first application of X-ray investigation to inorganic crystalline substances, to their use in the study of the more complex organic compounds associated with living organisms and finally of the protoplasm itself. The author stresses the fact that not only in non-living organic materials, but also in protoplasm a crystalline pattern has been found. The difference between inorganic and organic substances would lie then in the much greater complexity of the latter rather than in a different method of construction.

Increased knowledge and new methods of research have also produced in recent years a large body of data on the size of a wide range of proteins and viruses, and comparisons have been made with the size of cellular units of living organisms. In this respect a continuum can be found from the smallest ultramicroscopic protein molecule to the viruses, bacteria, and finally to cellular units of microscopic visibility. Viruses themselves form an unbroken series, with regard to size, from protein molecules to bacteria, with some overlapping at either end.

Speculations based on this kind of evidence induce the author to ask whether the cell, as a structural unit, has to be considered "an entirely distinct and unrelated entity in the material world or does it represent the climax of a gradually increasing complexity in pattern having direct continuity with the structural forms of less complexity associated with the materials of the non-living world?"

The controversial issue as to whether viruses are enormous protein molecules or elementary living units is brought up. The author points out the apparent correlation between complexity of structure and size on one hand and function on the other. Certain of the features associated with function at the cellular level are lacking in viruses, but others are present. This again represents a step in a series where complexity of structure is associated with higher function.

The intricate organization of the substance forming the cell is outlined. An analogy is sought between the association of the protein molecules in the protoplasm of each cell and the association of "cell-molecules" in the multicellular organisms. Throughout the paper a unifying principle is invoked to explain all types of organization, whether in the inorganic realm or in the organic. The question of the identification of the ultimate and indivisible unit of life is left open.

2 *Bergmann, G.* **Reduction.** In *Current trends in psychology and the behavioral sciences.* Pittsburgh: University of Pittsburgh Press, 1954.

Every theory possesses (*a*) a class of undefined or basic terms, (*b*) a class of terms derived from the basic by appropriate definitions, (*c*) a class of statements called axioms (presumably assumed), and (*d*) a class of statements called theorems deduced from the axioms.

Starting with geometry as the oldest and most familiar example, Bergmann notes that the theories of all sciences have introduced the time dimension (associated with "process") which geometry lacked. It became the task of theory not merely to deduce a full description of a configuration from a partial description, but also to predict future states (or infer past states) of systems from observed states. A theory is called comprehensive if (*a*) all variables of the area are definable in terms of the undefined variables (notions, terms) and (*b*) all laws which con-

tain only the variables of the area are theorems of the theory.

Reduction involves a certain relation between two theories. The theory to be reduced is called a makrotheory; the one to which it is reduced, a mikrotheory. Reduction involves assigning a correspondence between (*a*) each term of the makrotheory and a derived (not basic) term of the mikrotheory and (*b*) each statement of the makrotheory and a theorem of the mikrotheory.

Four "classical" cases are cited of which only the first is evaluated as completely successful. They are—presumably in a decreasing degree of success—the reduction of (*a*) gas theory to mechanics, (*b*) physiology to physics-chemistry, (*c*) psychology to physiology, and (*d*) social sciences to psychology. Aside from the degree of success achieved, these instances of reduction differ in other important respects. In the first instance, the basic entities of the mikrotheory are the refined constructs rather than the observed objects. The correspondences between gas theory and mechanics are therefore *made*, not discovered; one searches mentally for such correspondences until one finds some that fit. In the second and third instances, however, the correspondences (translation of physiological to physico-chemical laws or from behavior to physiological correlates) were factually *discovered*. In the last instance, reduction is predominantly a problem of definition, for example, the problem of translating group terms (e.g., rate of growth of per capita income) or humanistically intuitive ones (e.g., the ideology of the church) into the proper terms of the mikrotheory.

The role of reduction in providing the cement to bind the isolated relations of "incomplete knowledge" (empirical regularities) is discussed. The demand for reduction, however, as a prerequisite to promoting a theory to scientific status is dismissed, especially on the ground of the inadequacy of the arguments usually supporting such demands.

Finally, a distinction is made between systematic and historical process theory. The first is governed by the assumption that it is possible to predict and postdict states of a system on the basis of "momentary" observations; the second necessitates the knowledge of both a state and a history leading to it. The attempts of the Gestaltists to bridge the gap by postulating "traces" (to account for the history at the moment of observation) is taken into account but only as a "promissory note." Physical and physiological theories are cited as systematic; psychology as primarily historical; social science theory as systematic at least in intent. The question whether systematic laws can be deduced from historical ones and therefore whether one type of theory can be reduced to another is left open.

3 *Fitts, P. M.* **The influence of response coding on performance in motor tasks.** In *Current trends in information theory*, Pittsburgh: University of Pittsburgh Press, 1953. Pp. 47–75.

Human motor system's capacity in generating information is discussed in terms of reaction time and accuracy.

Decision time, as a function of the number and kinds of alternatives from which a motor response is selected, is analyzed. The results obtained in experiments dealing with the selection of a particular response member and others dealing with the selection of the direction or amplitude of the movement of a designated body member, are compared. Under the name of "Binary Index of Task Performance," a unit of measurement is proposed that will permit the comparison of performance rates under different amplitude and accuracy conditions. Measured in these units, the rate of motor performance is relatively uniform over a wide range of conditions in a given task and is somewhat comparable in different tasks.

The importance of the "matching" of stimulus coding and response coding is emphasized. If there is no "compatibility" between stimulus and response, the information present in the stimulus must be transformed or re-encoded before the motor response can take place. Each re-encoding step would add an additional time delay and increase the probability of an error. The fact that in man the fastest decisions have been found to be the most accurate is consistent with the hypothesis that transfer of information from stimulus to response is in such cases most direct, i.e., requires a minimum of re-encoding.

4 *Frick, F. C.* **Some perceptual problems from the point of view of information theory.** In *Current trends in information theory*. Pittsburgh: University of Pittsburgh Press, 1953. Pp. 76–91.

The general process of perception is considered in terms of information theory. The observer is thought of as a communication channel, and his channel capacity, or the amount of information that he can transmit, is discussed.

On the basis of some simple observations, it is evident that the order of magnitude of such capacity compares unfavorably with that of a radio or television channel. However, more insight is gained in this discrepancy if some of the discriminative restrictions which limit human channel capacities are considered.

Starting from some of Pollack's experiments, Frick points out that the observer's ability to discriminate depends upon the ratio "of the difference between stimuli within a series to the total range of permissible, or probable, variation" (signal-to-noise ratio). Experiments performed by Frick and his co-workers seem to lead to the conclusion that the signal-to-noise ratio for perfect discrimination is relatively invariant for a number of different stimulus aspects. On the basis of another series of experiments, he concludes that the recognition time for a set of printed words increases linearly with the number of responses available, i.e., with the amount of information needed by the observer.

Even if accurate identification depends, in large degree, upon the amount of information needed and is relatively independent of the specific stimulus characteristic, there are occasions on which the situation does not seem to furnish the *kind* of information needed to make an appropriate decision. An analogy can be made here with what happens with a filter when the input is not perfectly matched to the filter characteristic. Therefore, in order efficiently to encode information for transmission through the human channel, a proper match between the stimulus input and the response output has to be insured.

5 *Jacobson, Homer.* **Information, reproduction and the origin of life.** *Amer. Scientist*, 1955, 43, 119–127.

This paper deals with the informational aspects of reproduction and the origin of life. The consideration of living organisms as sources of information has brought to light several fruitful new ideas in these important fields: (1) The idea of a complexity measure, applicable equally to zygotes, spores, mature organisms or mechanical systems; (2) the use of a complexity measure to estimate the minimal complexity of the first living organism; (3) the use of a complexity measure to estimate the minimal complexity for self-reproducing devices; and, (4) a comparison of the estimates of minimal complexity, stated in (2) and (3), to investigate the possibility for the spontaneous origin of a living, self-reproducing species.

An informational approach as just outlined, focuses attention on the functional qualities of life systems, rather than on the minutiae of their structure. These functional examinations suggest that the reproduction of chromosomal material in a cell nucleus may be a wholly distinct function from the duplication of the cytoplasm; and that the self-duplication phase of reproduction is primarily cytoplasmic, rather than chromosomal. (Author's summary.)

A GAME THEORETIC ANALYSIS OF CONGRESSIONAL POWER DISTRIBUTIONS FOR A STABLE TWO-PARTY SYSTEM[1]

by R. Duncan Luce and Arnold A. Rogow

Columbia University and State University of Iowa

Attempts to make behavioral science a science often hinge on making vague terms precise or quantifiable, if possible. From game theory comes a definition of "power" (as applied to coalitions) in precise and quantifiable form. Once one has a quantitative definition of power, one wonders how this quantity will be distributed in a body of men with partially the same and partially opposite interests and under the influence of various other considerations. It is assumed here that the American legislative apparatus (involving both Congress and the President) is such a body. The picture is necessarily a drastically simplified one, but it has some essential and recognizable features of its referent: majority rule, the veto, the overriding of the veto, the two-party system, the defections of some members from one party to the other, the "die hards" who never defect. The results of rigorous analysis are in general in agreement with the "findings" of the political scientists about where the power resides under various conditions. But in addition this game-theoretical approach calls attention to other results, which are not often emphasized or realized. The method thus provides a framework for future enrichment and refinement of exact methods in political science.

THE theory of games (11) is a mathematical model for conflict of interest among intelligent and motivated agents; it is, therefore, not surprising that attempts are being made to apply it to some political science problems. So far, there are two such published efforts. The first consists of readings in game theory, assembled by Shubik (9), which serve to introduce some of the basic ideas and criticisms of the theory. The second, a paper by Shapley and Shubik (8), is the first actual application of a portion of game theory to a political science problem. The present paper outlines another application, which is in a sense a natural successor to the Shapley-Shubik work.

We must emphasize that this is an outline, for we do not view the content as a serious attempt to study Congress as such; rather, it is a purposely oversimplified illustration of how a part of game theory may be applied to such studies.

To this end, we have taken pains to point out the nature of the assumptions made, the extent to which they are inherent in the model, and the extent to which they are simply matters of convenient exposition. We hold that a much more serious attempt must be made to abstract the central features of a legislature than we have undertaken before one can decide whether the particular model recommended is useful or not. In apparent

[1] This work was undertaken when the authors were Fellows (1954–55) at the Center for Advanced Study in the Behavioral Sciences, Stanford, California. The final revision was supported in part through funds extended to the Behavioral Models Project, Bureau of Applied Social Research, Columbia University by the Office of Naval Research. The paper may be identified as publication A198 of the Bureau of Applied Social Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

contrast to our doubts that the formulation will sustain careful scrutiny by students of Congress, we have formulated a group of generalizations that summarize some of our detailed results and we have discussed these in the light of prevailing generalizations and some data. This should not be misinterpreted; it is only intended to demonstrate how it is possible to go from the mass of detail generated by the model to the types of generalization more familiar to a political scientist. We hope that our discussion will stimulate others to use these mathematical tools in deeper analyses of legislative bodies —studies in which data are related more closely to theory than we have done.

A legislature, as a voting body, can be viewed as having two inherently different aspects: a formal body of rules, called the legislative scheme, which establish the conditions under which a bill is passed, and the various peculiarities and limitations characteristic of a particular legislature working within the given scheme. Included in these limitations are such realities as the party structure, party discipline, the effects of pressure groups, etc. Following Shapley and Shubik, it appears that the scheme is usefully identified with one of the central notions in n-person game theory: the characteristic function. It is suggested that many of the "realities" of a legislature can be identified with another notion central to one of the several equilibrium theories. This theory is concerned with those couplings of a "power distribution" to a division of the participants into coalitions such that no changes occur. Since it is striking that the two-party system—a division of the participants into two disjoint coalitions—has remained stable for a long time, we shall assume that it *is* stable and inquire as to the theoretically necessary location of power in Congress and the presidency for such stability. We shall not attempt to discuss the much more profound question as to why the two-party system has evolved and why it is stable, but only what conditions are theoretically necessary on the power distributions in order that the system be stable. As will be seen, the theory is simply a formalization of the usual verbal discussions about the location of power—a formalization that can be readily extended to more complicated models of Congress where many of us would find it difficult or impossible to extend a nonsymbolic analysis.

Our work relates to the Shapley-Shubik paper in the following way: They ignored any information one might have about specific legislatures and confined their attention to legislative schemes. Given such a scheme, they inquired into the possibility of a priori statements about the power distributions implied by the scheme. If one is willing to accept certain conditions as to the nature of such an a priori distribution, they establish that it is uniquely determined by the voting scheme. For the study of existing legislatures this is obviously insufficient and one must take into account some of the known realities and attempt to deduce from them other—known or unknown—assertions and to investigate the empirical truth of these consequences. Our purpose is to begin to deal with this problem.

Before turning to the details of the model, let us freely admit that to many—particularly to those with mathematical training—the following discussion will seem tedious and the content slight in relation to the length. It is true that the content could be contained in half a dozen or eight pages, but only at the expense of using more mathematics and thereby excluding as readers many of those we most want to address: political scientists concerned with the study of legislatures. One price of interdisciplinary communication seems to be length.

1. A mathematical representation of a legislative scheme

For the purposes of this paper, we shall suppose that the only function of a legislative body is to vote on bills which are presented to it; of course, this is not actually the case, but it may prove to be a suitable first approximation to a legislature and, at the least, it should be of interest to see what consequences can be drawn from it. For the moment, we shall not consider how a legislature may be divided into chambers nor how it may be partitioned into parties;

rather, we will simply think of it as a body of undifferentiated men who vote. Let us denote this set (the terms "class" or "collection" are also used) of men by the symbol L (which stands for "legislators"). Consider any subset S of the set L (for example, if L denotes the set of men in the United States Congress on January 1, 1955, then the Southern Democratic Senators form a well-defined subset of L). The rules of the legislative scheme under consideration must determine whether or not this set S is able as a voting coalition to pass a bill. Indeed, one way to prescribe the voting rules of the body is to list for each possible subset S whether or not it can pass a bill. Let us call those coalitions which can pass a bill *winning* and those which cannot *losing*.

The legislative scheme with which we shall be concerned is that of the United States. It involves two sets of men: Congress, which we shall denote by the symbol C, and the President, whom we shall denote by P. For our purpose, it will not prove necessary to take into account that Congress is divided into two houses, for we shall not be concerned here with the origination of bills, with committee activities, or with treaties, and we shall always assume similar majorities in both houses. A coalition in this scheme is winning if and only if either

1. it consists of a majority (in both houses) of Congress and the President, or
2. it consists of a two-thirds majority (in both houses) of Congress.

All other possible coalitions are losing. It should be noted that by ignoring the possibility of ties, we have made a minor idealization: this is not essential to the model and it can be eliminated at the cost of more *routine* labor later on.

Returning to the general case, we may assume that as a result of passing or defeating a bill, there are certain rewards accruing to those involved. These may range from outright money payments to individuals, through various forms of indirect compensation, to changes in relative prestige. Each of the individuals in the legislature is assumed to have a pattern of preferences among these outcomes. While it is very difficult to ascertain these preferences in practice, they may still be postulated. If they are defined over all risky outcomes, i.e., outcomes consisting of chance distributions over the basic outcomes, and if they satisfy certain axioms, then the theory of utility establishes how the preferences may be represented by numbers (11, pp. 15–30). Since there is an extensive literature on this subject, brevity dictates that we cannot delve into it deeply. Nonetheless, it must be emphasized that the theory is controversial and that many authors do not feel that people can be expected to exhibit the consistency demand by the axioms. On the other hand, the axioms do have a certain compelling plausibility. One important difficulty in the theory, as it is now developed, is that the numbers representing preferences are not uniquely determined—the choice of both the zero and the scale unit is arbitrary. The important ambiguity is that of the unit, for it is impossible to say what changes in the underlying outcomes result in the same utility change for two different people. This is the famed problem of interpersonal comparisons of utility.

Assuming the existence of such utility functions and a solution to the problem of interpersonal comparisons, the theory of games shows how it is possible to derive a number for each logically possible coalition which represents, in utility units, the amount of the rewards that the coalition as a whole may insure for itself (11, pp. 238–243). The fact that interpersonal comparisons cannot now be made would seem to render this construction empty; however, it is not completely empty if one supposes there exists a solution to this problem and, in some contexts, it is possible to determine these summary numbers without actually obtaining the individual utilities (see below). In a sense, such a number associated with a coalition represents its power with respect to the single bill under consideration. When we use the word "power" in the rest of this paper, it shall mean only the numerical representation of rewards accruing to coalitions as evaluated by the members of these coalitions. It is important that no other

meanings of this word be read into our results.

The power, in the sense of the numbers just described, of a coalition depends not only on the ability of the coalition to pass a bill, but also on the individual evaluation of the outcomes; thus, we cannot anticipate in general that these numbers will meet very strong restrictions. In fact, in the theory of games they are shown to satisfy only two very reasonable requirements; these are discussed in the Appendix. Since our work is primarily illustrative of a mathematical method, we shall not attempt to deal with the completely general case, but rather we shall suppose that the individual evaluations of the rewards resulting from passing the bill under consideration are such that each of the winning coalitions has identical power and each of the losing coalitions has identical power. This, very clearly, is an idealization, but one which may be approximately true in some cases. It is not, however, essential to the methods we are illustrating, as is pointed out in the Appendix, so long as one can devise empirical methods for estimating the relative power of the various coalitions. As we shall see, it will not prove necessary to have these estimates for all possible coalitions, but only for a relatively few relevant ones.

With the assumption of the power equality of all winning coalitions and of all losing ones, there is no loss of generality in setting the power of a winning coalition equal to 1 and that of a losing one equal to 0; this we shall do. If S is any subset of L, we denote by $v(S)$ the power of S acting as a coalition, i.e., $v(S) = 1$ if S is winning and $=0$ if S is losing.

Of course, it is a theoretical fiction to speak of the power of a coalition. True, the power results from the collection of men acting as a coalition, and so in that sense it is associated with the coalition, but the rewards it represents must actually be rewards to individuals in the legislature. We cannot even say "rewards to just the members of the coalition which passes the bill," for the coalition may find it expedient to turn over some of the rewards to men not in the coalition. At least this is an a priori possibility and though it will not actually occur under the assumption of equality of power for winning and for losing coalitions, with more general assumptions it can. We, therefore, suppose that the distribution of total power to the different legislators can be represented by a collection of n numbers x_i, where i is an index running over the n legislators. For example, if we number the legislators from 1 to n, then x_{10} denotes the power accruing to the legislator numbered 10. We shall stipulate that *all* the power is distributed to the legislators, i.e.,

$$x_1 + x_2 + \cdots + x_n = 1,$$

and that the smallest amount of power is 0, i.e.,

$$x_i \geq 0 \qquad \text{for } i = 1, 2, \cdots n.$$

(In the vocabulary of the theory of games, such a distribution is called an imputation.)

We shall suppose, subject to some limitations to be given later, that during the prevote haggling and threatening each of the legislators is attempting to achieve as large a portion of this distribution of power as he can. The purpose of our analysis, among other things, will be to establish which, if any, of the distributions of power are in equilibrium in the sense that further haggling will not result in a modification of the distribution.

At this point, it is convenient to interrupt our pursuits to indicate what Shapley and Shubik have done in the paper mentioned earlier (8). Given a legislative *scheme*, as described by the winning and losing coalitions, they inquired what, if any, a priori statements could be made about power distributions. They quite consciously ignored all the specific information one might have or might obtain about a specific legislature, such as Congress at a certain date, and they concentrated entirely on the information given by the legislative scheme. We shall not attempt to reconstruct their argument, but we may mention the nature of the important and surprising theorem of Shapley's upon which their work rests. He has shown (7) that if an a priori distribution of power is required to meet some (apparently) weak and possibly acceptable conditions, then

there is a unique answer to the problem which can be expressed by a simple formula involving the numbers $v(S)$. The formula amounts to calculating for each legislator the chance that, if a coalition were built up successively by random selections from among the legislators, he would be the individual who converted it from a losing one to a winning one.

They presented several calculations of a priori power distributions for well-known voting schemes. For example, for the congressional-presidential scheme, the ratio of power of a single representative to the presidency is 2 to 350 and between a senator and the presidency the ratio is 9 to 350. If we take the House of Representatives and the Senate as single units, then they have equal a priori power and the presidency is two-fifths as powerful. If the congressional scheme were modified so that the presidential veto could not be overridden, then the House would have slightly less power than the Senate and the presidency would be about twice as powerful as either of them. As they point out, such a theory gives one a tool to examine the effect of revising legislative procedure, for "... it can easily happen that the mathematical structure of a voting system conceals a bias in power distribution unsuspected and unintended by the authors of the revision" (8).

Our purpose is to go beyond the Shapley-Shubik analysis of a legislative scheme to an analysis of a legislature. This will, of course, necessitate a model of what we mean by a legislature. In the next section, we shall present a model for Congress, which is illustrative of a class of models for legislatures. (The general class is discussed in the Appendix.) These models must attempt to capture some of the realities of specific legislatures—realities which are not part of the voting scheme. In the case of Congress, we mean by realities such facts as the party structure, the committee roles, the liberal-conservative dichotomy, the individual loyalties to party, personal animosities, etc. We shall not, by any means, attempt to deal with all of these in this illustrative example, but only with the party structure and an approximation to party loyalty. This will permit a plausible first approximation which is sufficiently simple to render the analysis fairly transparent. The effect of introducing other factors is only to increase the details of analysis without modifying the basic procedure.

2. A model of Congress

We shall start with the fact that every legislator is identified with one of two non-overlapping political parties. We shall assume for simplicity, and with little practical loss of generality, that whichever party has a majority in one chamber of Congress has a majority in the other. Let us label the majority party as number 1 and the minority party as number 2, and let us denote by C_1 the set of congressmen in the majority party and by C_2 the set in the minority party.

An arresting fact about Congress is the stability of the two-party system, i.e., the simple fact that it has not split into more than two parties or reduced to one. One of the major questions to which we shall address ourselves is whether there exist distributions of power for our model of Congress which permit a stable two-party system. Thus, if the President is a member of the majority party, we shall be interested in the stability (in a sense yet to be defined) of the partition of the voting body (which includes the President) into C_2 on the one hand and C_1 plus P on the other. Let the set consisting of C_1 and P be denoted by $C_1 \cup P$. Let τ_1 denote this partition, $(C_1 \cup P, C_2)$. Similarly, if the President is in the minority party, we shall be interested in the partition

$$\tau_2 = (C_1, C_2 \cup P).$$

While the two-party structure is known to be stable, it is equally clear that some, if not most, bills are passed by coalitions different from the party coalition. The conservatives of the two parties may join as a temporary coalition to pass a single bill without causing the disruption of the party structure. However, given a particular bill and a particular Congress, there are certain coalitions which could not conceivably form.

If we restrict our attention to the partitioning of Congress by parties, these limitations on the formation of coalitions (to pass a bill) are, therefore, limitations on defections from the parties. Such limitations are produced by a number of factors such as pressure from constituents, party discipline, pressure from lobbies, the particular issue at stake, etc. However they may be generated, they can be described in terms of the set of Congressmen who can be induced to defect to the other party. Here we make two simplifying assumptions: first, that there are no abstentions from voting and, second, that the only defections are from one party to the other. Actually, the second of these omits the important possibility of defectors from both parties joining to form a winning coalition. Again, as will be seen, there is no inherent reason for making this assumption: it only reduces the amount of routine calculation and the amount of space needed to present the results.

Because of the nature of the legislative scheme, there are only two groups of defections which are of interest: a defection which swells the ranks of the other party to a two-thirds majority in each house, in which case a presidential veto can be overridden, and a defection which fails to achieve a two-thirds majority in at least one house but does result in a simple majority in each, in which case a coalition of the majority and the President can pass the bill. In either case, the party in question theoretically will only be interested in defections which just produce the desired majority—all other votes are technically superfluous. Now, at the time of any given vote, it does not seem too implausible to suppose that in principle the potential defectors in each party can be graded from the most to least willing to defect from their party. If so, and if the defectors are added to a coalition in order of decreasing willingness to defect, then there is a unique set of defectors which will just create a simple majority, if that is possible, and a unique set which will just create a two-thirds majority, if that is possible. So, when we speak of a set of defectors which create a certain majority, we shall mean that minimal set of congressmen, drawn from among those most willing to defect, which is just necessary to form the majority.

With this assumption, then, we may divide the congressmen of parties 1 and 2 into two non-overlapping sets: C_1' and C_2' will denote the sets of *potential defectors* from parties 1 and 2 respectively and C_1'' and C_2'' the remaining members of each party, who will be called the *diehards*.

There are a number of different cases which can occur in this idealized Congress. Since there is no a priori reason to exclude any of these we shall examine each of them separately, and on the basis of this exhaustive examination, we shall draw some general qualitative conclusions from the model (section 5).

First, there are four basically different partitions of Congress into the two-party system: either party 1 (which by assumption has a majority) fails to have a two-thirds majority in at least one chamber, or it has a two-thirds majority in both, and either the President is a member of party 1 or of party 2.

Second, there are twelve different possible limitations on coalition changes from the two-party partition. These twelve arise from a selection of one of the alternatives in each of the following three classes of alternatives:

1. either the President is (or feels) free to defect from his party, or he is not;
2. party 1 plus the defectors from party 2 either form only a simple majority, or they form a two-thirds majority in both houses; and
3. party 2 plus the defectors from party 1 either fail to form a simple majority in at least one house, or they form a simple majority in both houses but not a two-thirds majority in at least one, or they form a two-thirds majority in both houses.

We observe that not all these limitations are compatible with all the coalition partitions, e.g., it is not possible for the majority party to have a two-thirds majority in both houses and for the addition of defectors from the other party to reduce this to a

simple majority. If such cases are excluded, then there are a total of 36 cases to be examined.

3. Equilibrium power distributions in a special case

To introduce the equilibrium notion we shall use and to illustrate the typical calculations involved, we shall choose one of the 36 cases; it matters little which one we take. Let us suppose the President is in party 1 and that party 1 has only a simple majority in Congress. Thus, we shall be concerned with the partition $\tau_1 = (C_1 \cup P, C_2)$, where C_1 does not form a two-thirds majority. As the system of limitations on changes from τ_1, let us choose the case where

1. The President is free to defect;
2. party 1 plus the defectors from party 2 form only a simple majority; and
3. party 2 plus the defectors from party 1 form only a simple majority.

Now, from these assumptions, we see that a two-thirds majority does not exist among the possible changes; thus any winning coalition that can form must include the President. It is easy to see that there are only three such coalitions:

$$C_1 \cup P,\ C_1 \cup C_2' \cup P, \text{ and } C_2 \cup C_1' \cup P,$$

where, as before, if A and B are sets, $A \cup B$ is the set whose elements consist of those in A and those in B. One of these three will have to form to pass the bill; we shall not concern ourselves with which (there are certainly not enough assumptions even to begin to answer that question), but rather with the location of power in such a Congress under the assumption that the two-party system is stable.

Consider a power distribution x_i, $i = 1, 2 \cdots n$, which is offered as being compatible in this Congress with the two-party partitioning of Congress. If any legislator i not in $C_1 \cup P$ has $x_i > 0$, then we know that the sum of the x_j for j in $C_1 \cup P$, must be less than 1. But since $C_1 \cup P$ is winning it can command power of 1, and so it can form and each of its members can receive more than they did in the arrangement x_i, $i = 1, 2, \cdots n$. Since we are assuming each legislator is out to better his gains of power, we must conclude that if anyone outside $C_1 \cup P$ has power, then the combination of the two-party partition and the given power distribution cannot be in equilibrium. A similar argument applies to $C_1 \cup C_2' \cup P$ and to $C_2 \cup C_1' \cup P$, but not to any other set of legislators for either they are losing, and so can offer no power gains, or they are not among the admissible changes. Thus a power distribution x_i, $i = 1, 2 \cdots n$, is in equilibrium with τ_1 only if $x_i = 0$ for any legislator i not in *each* of the three coalitions. Thus, the power must be distributed over those who are in *all* three of the winning coalitions, i.e., over $C_1' \cup P$. In other words, in this situation the power is distributed over the defectors from party 1 and the President. Exactly how it is distributed over these men is not determined in this simple model: it presumably depends upon factors which we have not taken into account. If we were to extend the model to include such things as the actions of committees, the origins of the bill, and so on, we could then expect a much more detailed determination of the power distributions.

4. The equilibrium power distributions for the two-party system

In order that there shall be no ambiguity in our definition of equilibrium, we shall be somewhat more formal than we have been up to now. Let $v(S)$ denote the number representing the power accruing to the set S if it forms a voting coalition. Let X stand for the distribution of power x_i, $i = 1, 2, \cdots n$. Let τ denote a partitioning of the legislature into coalitions (while we have assumed partitions into only two coalitions, this assumption is by no means necessary). Let the symbol $\psi(\tau)$ stand for the class of coalitions which can form, if there is any reason to do so, when the legislature is partitioned according to τ. Then, we shall say that the pair (X, τ) is *ψ-stable* (which is simply the name assigned to this particular definition of equilibrium) if for each set S in $\psi(\tau)$.

$$v(S) \leq \sum_{i \text{ in } S} x_i .$$

(The last symbol, $\Sigma_{i \in S} x_i$, simply means the sum of the values x_i for each legislator i in the coalition S.)

A re-examination of the analysis of the special case discussed above shows that it is just a special case of this definition.

Actually, the definition we have given here is slightly different from that presented in the mathematical literature (3, 4). There it was stipulated that if a legislator receives no power at all, i.e., if $x_i = 0$, then he shall not be involved in a coalition with any other legislators. The argument for imposing this condition arises simply from asking why he should cooperate with others if this cooperation does not result in any gain for him. This is a powerful argument in situations which occur only once, but it seems less convincing in a legislature where one bill is but part of an on-going process.

Now, while we have in fact been ignoring the on-going process (see the Appendix),

TABLE 1

THE POWER DISTRIBUTIONS FOR A STABLE TWO-PARTY SYSTEM UNDER THE GIVEN CONDITIONS

No.	Presidential Defection	Party of President	Size of Party 1 Majority	Size of Party 1 plus Party 2 defectors	Size of Party 2 plus Party 1 defectors	Locations of Power
1.a.i	Possible	Either	simple	simple	less than majority	Party 1, President
ii					simple	Party 1 defectors, President
iii					two-thirds	Party 1 defectors
1.b.i				two-thirds	less than majority	Party 1
ii					simple	Party 1 defectors
iii					two-thirds	Party 1 defectors
2.a			two-thirds	two-thirds	less than majority	Party 1
b					simple	Party 1 defectors
c					two-thirds	Party 1 defectors
3.a.i	Not Possible	1	simple	simple	less than majority	Party 1, President
ii					simple	Party 1, President
iii					two-thirds	Party 1 defectors
3.b.i				two-thirds	less than majority	Party 1
ii					simple	Party 1
iii					two-thirds	Party 1 defectors
4.a.i		2	simple	simple	less than majority	Impass-no winning coalitions
ii					simple	Party 2, Party 1 defectors, President
iii					two-thirds	Party 2, Party 1 defectors
4.b.i				two-thirds	less than majority	Party 1, Party 2 defectors
ii					simple	Party 1 defectors, Party 2 defectors
iii					two-thirds	Party 1 defectors, Party 2 defectors
5.a		1	two-thirds	two-thirds	less than majority	Party 1
b					simple	Party 1
c					two-thirds	Party 1 defectors
6.a		2	two-thirds	two-thirds	less than majority	Party 1
b					simple	Party 1 defectors
c					two-thirds	Party 1 defectors

we shall at least take it into account to the extent of waiving the second condition. It may also be worth noting that from the mathematical point of view this does not constitute a very serious change in the definition for the second condition has not played an important role in most of the theorems proved.

We may now give (Table 1) the power distributions which coupled with the party partitioning of Congress are stable in each of the 36 cases described in Section 2. The conclusion in each case arises by an argument like that given in Section 3 (that case is number 1.a.ii in Table 1).

5. Conclusions

By examining the summary of results in Table 1, certain generalizations implied by the model are clear. We may list six which seem interesting:

1. In all cases the arrangement of Congress into two opposed party coalitions is stable provided the power is distributed as indicated. In very many cases, however, it is necessary to form coalitions other than along party lines in order to produce a winning coalition, i.e., to pass a bill. In only one case (4.a.i) are the limitations so stringent that no working majority can form: this is when the President is of the minority party and will not defect to the majority, the majority party has only a simple majority even with the defectors from the minority, and the minority does not have a simple majority even with the defectors from the majority. What is interesting is that in only one case of the 36 can such an impasse result.

2. In all circumstances, the President is weak when the majority party—whether he is a member of it or not—has a two-thirds majority. If this model has any relation to reality, we must conclude that a President should fear a real Congressional landslide for *either* party.

3. The President possesses power (from voting considerations) only when neither party can muster more than a simple majority even with the help of the defectors from the other party.

4. The only circumstances when the minority party is the holder of any power is when the President is in the minority party and he is unwilling to defect to the majority.

5. Under all conditions, if the defectors from party 1 added to party 2 fail to form a majority, then the diehards of party 1 possess power. The only other case in which they possess power is when the President is a member of party 1, he is unwilling to defect, and party 2 plus the defectors from party 1 form only a simple majority (cases 3.a.ii, 3.b.ii, and 5.b).

6. The only case when the party 2 diehards possess any power is when the President is a member of their party, he is unwilling to defect, party 1 has only a simple majority, and party 2 plus the defectors from party 1 form either a majority or a two-thirds majority (cases 4.a.ii and iii).

In connection with these last two statements, we note that there are a large number of situations where the diehards are not in a position to command power in a stable two-party system, and those situations in which they do have power appear to be ones not likely to occur often in practice. Recall that when we introduced the concept of a ψ-stable pair (section 4) we noted that in the original mathematical definition it was stipulated that a nontrivial coalition would not exist if its members did not benefit from their participation in the coalition. While we waived this condition for our work in this paper, there is still some force to the argument if we think of the long-term existence of a legislature. Thus, if voting coalitions tend to stabilize in time and if consideration is not given to the diehards by the other party members, it should not be too surprising to find a sudden change in their behavior. Of course, it must be understood that this is not a conclusion from the model but an extrapolation, for the model (as a special case of game theory) is inherently static and does not attempt to deal with such changes in time.

Despite the various limitations of the model, which we have already noted, the generalizations we have drawn from it do not appear to be seriously inconsistent with a number of political science "findings"

concerning Congress, and in several cases they appear to emphasize aspects of congressional power which may not have been given adequate attention. The sorts of statements with which the model is not in serious contradiction are:

The President, as a legislator, is weak when either

1. his party is in the minority;
2. his party is in the majority but is not committed to his program; or
3. either party can muster sufficient strength to overturn a veto:

and the President, as a legislator, is strong when either

4. his party is in the majority and is committed to his program;
5. his party is in the majority, whether or not it is committed to his program, provided that it cannot muster sufficient strength to overturn a veto;
6. supporters from his own party and defectors from the opposition party constitute a pro-administration majority coalition.

Such statements as these are not in all cases fully acceptable generalizations from the model, e.g., the first one must be qualified by the condition that one or another congressional coalition has a two-thirds majority, for if not the President possibly does possess power (see cases 1.a.i, 1.a.ii, 3.a.i, 3.a.ii, and 4.a.ii). A careful examination of our results will show that several others of these statements need some qualification before one can say that they coincide with the results of the model, but the spirit of them is much the same. It would be of some interest to know how stable these generalizations are under slight changes in the assumed limitations on coalition changes (as given by $\psi(\tau)$): for example, one might examine the effect of including as possible coalitions those made up from the defectors of the two parties. This would result in many more cases to examine, but none would be any more difficult to deal with than those above.

Let us emphasize once again that both these statements and the generalizations arrived at using game theory refer only to one aspect of Executive-Legislative relations. They do not deal with the crucial power position of committees and party leaders, the special position of the Rules Committee in the House, the filibuster power in the Senate, or a number of other important factors. Consequently, no claim can be made that our generalizations would not be substantially modified were the model refined.

On the other hand, they do emphasize several factors which are sometimes neglected or which are occasionally minimized in political science literature. One of these is concerned with the power position of diehards vis-à-vis defectors. As V. O. Key, Jr.[2] and James M. Burns, (1, see also 2) among others, have noted, party defectors often constitute either the effective working majority or the effective opposition, irrespective of whether or not they are members of the nominal majority party. The model suggests that in only a relatively few (and relatively unlikely) cases do the party diehards possess effective power. In the majority of the 36 cases, party defectors hold the balance of power.

This observation suggests that more attention needs to be paid to the analysis of Congress not in terms of nominal party majorities or minorities but in terms of cross-party groupings which might be tentatively classified as we have done. Thus far, empirical research does not provide a clear answer to the question of the extent of party cohesion,[3] let alone answers to

[2] "In the American Congress the weak ties of the majority encourage the minority to wean away followers from the majority party and to determine the outcome of at least some and at times many legislative issues. It is not uncommon for the working 'majority' to be composed of a substantial part of the minority plus a sector of the nominal majority . . . the genuinely effective 'opposition' often consists, not of the minority, but of recalcitrant members of the majority who hold a balance of power within the House or Senate" (6, p. 706).

[3] The failure is reflected in the differing views held of the matter. Thus, Burns' view that "Party cohesion is still slight today," (1, 40) confirming A. Lawrence Lowell's analysis of fifty-four years ago, is similar to that advanced in the American Political Science Association's report "Toward a More Responsible Two-Party System." But a diametrically opposite position has been taken by a number of other political scientists. See for example, (10).

such questions as: who are the diehards and defectors, from what areas of the country do they come, why do they operate as such, and is the theoretically greater power of the defectors reflected in the legislation which finally gets to a vote, in committee assignments, or in other kinds of prestige and influence?

Our game theory generalizations also point up a number of inherent features of Executive-Legislative relations. It is clear from the model that the President, under the prevailing system of loose party discipline, does not necessarily gain when his party is returned to Congress with an overwhelming majority. Although most students of legislative behavior are of the view that the party of the President should have the main responsibility for the organization of Congress, our analysis suggests that the President is not much advantaged when his party elects two-thirds or more of the Senators and Representatives. Other things being equal, the President's power as a legislator appears to be great when his own party controls between 51% and 66% of Congressional seats, and when no cross-party combination of diehards and defectors can total more than two-thirds of the membership. Tactically, this means that his party should have only about 55% of the seats.[4]

Rather paradoxically, the President's power can also be great even when his own party is in the minority, provided that he is still in control of a bloc of his party possessing at least 35% of the Congressional votes. For in such cases, and so long as the 35% is loyal to the party chief, the majority (and opposition) party can initiate legislation, but it cannot subdue the veto power. The President, therefore, holds effective power as either party diehard or defector; in the former role, he can bring about a legislative stalemate, and in the latter role, he can transform an opposition which is impotent, so far as positive power is concerned, into a governmental party.

In closing, we must once again return to the limitations of the present version of the model. Whether or not this game theoretic analysis can provide us with any illumination of the Congressional power structure depends very largely on its potentialities of refinement to include factors which are known to be of importance. Thus, the question reduces to our ability to determine the power of coalitions, $v(S)$, and the limitations on coalition change, $\psi(\tau)$, for more complex real situations. More sophisticated models may very well necessitate the collection of empirical data in an attempt to determine these quantities. Presumably, the limitations on coalition defections, etc., can be assessed from obtainable data, in which case there seems to be no reason to assume, a priori, that the model must remain in its present elementary state. Equally, there seems no reason to suppose, if the generalizations deduced from the present model are at all interesting, that future refinements will not produce results of similar but more subtle interest.

[4] These generalizations are not contradicted by the record of Presidential vetoes cast during the Roosevelt Administration. As the following table shows, 69% of the vetoes cast by Roosevelt were cast during the three Congresses in which one or both houses were controlled by a two-thirds or better Democratic majority.

President	Year	Congress	House D	House R	Senate D	Senate R	Vetoes
Hoover	1931–33	72nd	220	214	47	48	18
Roosevelt	33–34	73rd	310	117	60	35	73
Roosevelt	35–36	74th	319	103	69	25	148
Roosevelt	37–38	75th	331	89	76	16	117
Roosevelt	39–40	76th	261	164	69	23	167
Roosevelt	41–42	77th	268	162	66	28	79
Roosevelt	43–44	78th	218	208	58	37	45
Roosevelt Truman	45–46	79th	242	190	56	38	76

Appendix

The purpose of this appendix is to indicate a little more explicitly the general form of the game model which we have illustrated by a special case in the main body of the paper. As we have indicated in several places, we believe that the methods discussed here are applicable to more refined models of legislatures; however, such applications surely will be more tedious in labor and will, in all probability, require the collection of empirical data.

As we said, game theory begins with a situation consisting of possible choices by individuals and individual outcomes associated to the choices. The individuals are assumed to have patterns of preferences among the outcomes which meet the axioms of utility theory, and so their preferences can be represented numerically. From this structure, the theory of games establishes that to each subset S of the individuals

there is assigned a number $v(S)$, which is calculated from the individual preferences. While in the model we examined $v(S)$ was either 0 or 1, in general this is not the case. If, for example, different rewards or different evaluations of the rewards occur for passing a bill, then the power of different coalitions which could pass the bill will be different. Or if we do not focus our attention on one bill, but on a series of bills, then since it is unreasonable to suppose that they will all result in the same rewards, we must suppose that the values $v(S)$ will depend upon more than the ability of the coalition S to pass a bill. The task of estimating $v(S)$ is clearly an empirical problem of some magnitude; there can be no doubt that it is a major stumbling block to the successful application of these methods. In particular, the theoretical construction of passing from individual preferences to $v(S)$ cannot be used, except in very special circumstances, to calculate $v(S)$; the practical difficulties are too great. More direct procedures are needed. One which may be useful are educated guesses made by well-informed observers. Another is a proposed, but untried, technique closely related to the methods of utility theory but applied directly to the coalitions and not to the individual outcome; this is discussed by Luce and Adams (5).

The collection of numbers $v(S)$, for all subsets S of the given set L (of legislators), form what in mathematics is called a real-valued set function v. This construct of game theory is called the *characteristic function* of the game. It might appear that since there is wide latitude in the individual evaluations of outcomes which lead to the characteristic function v, there can be no constraints on v in general. While this is almost true, it can be shown that it must meet two conditions, but in general no others can be established. Let us suppose that R and S are two subsets of L which have no legislators in common, and let $R \cup S$ be the subset consisting of the legislators in R plus those in S. Then it can be shown that

$$v(R \cup S) \geq v(R) + v(S).$$

In words, this simply means that the coalition $R \cup S$ can do everything R and S can do as separate coalitions, and possibly more. The whole is at least as great as the sum of its parts. The second condition, while mathematically significant, certainly appears trivial when one thinks of a legislature. It says the set ϕ which has no members—the empty set—has no power:

$$v(\phi) = 0$$

The interesting one is the former. In the case of the model we were discussing, it is particularly easy to see that it holds. If R and S are both losing $v(R) = 0$ and $v(S) = 0$. But $R \cup S$ may be either winning or losing, so the inequality holds. If R is winning and S is losing (or equally, R losing and S winning), then since $R \cup S$ is a set which includes the winning set R, it too must be winning. Thus, $v(R) = 1$, $v(S) = 0$ and $v(R \cup S) = 1$, and so equality exists. Finally, consider the case where both R and S are winning. Recall that we said R and S have no legislators in common, so we have the situation that if R passes a bill A which the coalition S does not like, then S can bring the negative of A before the legislature. Since S is winning it can pass not—A. Thus, the legislature would reach the deadlock where it could pass both A and not—A; clearly, no acceptable legislative scheme would have this property, so the assumption that both R and S are winning is not meaningful. Hence, we conclude that the inequality must be met.

As in the special model, the equilibrium states of the legislature, no matter what characteristic function is involved, are given in terms of a pair (X, τ), where X is a distribution of power to the individuals and τ is a partition of the legislature into coalitions. The distribution of power, or imputation as it is called, X must satisfy two conditions. It is a partitioning of the total power available:

$$x_1 + x_2 + \cdots + x_n = v(L).$$

and no individual receives less power than he can be certain of getting:

$$x_i \geq v(i) \quad \text{for} \quad i = 1, 2, \cdots n.$$

Finally, the model supposes that the realities of the situation, insofar as they produce limitations on shifts from one partitioning of the legislature to another, can be given in the following way. For each partition τ one is interested in, there is given a list of admissible coalition changes, which we denote by $\psi(\tau)$. It will be recalled that where τ was taken to be the partition into parties, $\psi(\tau)$ involved coalitions formed through defections from one party to the other. Again, it is an empirical problem to estimate these limitations. This problem will become particularly grave when an attempt is made to refine the model to take into account committees and other relevant features of Congress. Yet students of Congress have discussed such limitations and it may not be impossible to employ careful statistical studies to gain the necessary detailed data.

The application of the model is straightforward if the two empirical problems can be overcome. First we choose a partitioning τ of the legislature which we wish to investigate (possibly along party lines, but not necessarily). Second we determine $\psi(\tau)$. Third for each subset S in $\psi(\tau)$, we determine $v(S)$. Observe that so long as we deal with only one τ, it is not necessary to determine the whole of the characteristic function, but only the values for the admissible coalitions as given by $\psi(\tau)$. Fourth, we determine those distributions X which with τ are ψ-stable using the definition given in section 4. The fewer coalitions that are admissible, the less precisely defined is X, the more that are admissible, the more precisely defined is X. In many games, if too many coalitions are admissible, there is no X which meets all of the conditions of the definition of ψ-stability, in which case one is forced to conclude that the partition τ is not stable. The effect of the limitations given by ψ is to increase the stability of a legislature, or in general, of a game.

REFERENCES

1. Burns, J. M. *Congress on trial.* New York: Harper, 1949.
2. Carr, R. K., Bernstein, M. H., Morrison, D. H., Snyder, R. C., & McLean, J. E. *American democracy in theory and practice.* (Rev. ed.) New York: Rinehart, 1955.
3. Luce, R. D. A definition of stability for n-person games. *Ann. Math.*, 1954, 59, 357–366.
4. Luce, R. D. k-Stability of symmetric and of quota games. *Ann. Math.*, 1955, 62, 517–527.
5. Luce, R. D., & Adams, E. W. The determination of subjective characteristic functions in games with misperceived payoff functions. *Econometrica*, in press.
6. Key, V. O., Jr. *Politics, parties, and pressure groups.* (3rd Ed.) New York: Crowell, 1952.
7. Shapley, L. S. A value for n-person games. Contributions to the theory of games. II. *Annals of Mathematics Studies*, No. 28. Princeton, N. J.: Princeton University Press, 1953. Pp. 307–317.
8. Shapley, L. S., & Shubik, M. A method for evaluating the distribution of power in a committee system. *Amer. pol. sci. Rev.*, 1954, 48, 787–792.
9. Shubik, M. *Readings in game theory and political behavior.* New York: Doubleday, 1954.
10. Turner, J. Responsible parties: A dissent from the floor. *Amer. pol. sci. Rev.*, 1951, 45, 143–152.
11. von Neumann, J., & Morgenstern, O. *Theory of games and economic behavior.* Princeton, N. J.: Princeton University Press, 1947.

(Manuscript received October 22, 1955).

Laws and institutions must go hand in hand with the progress of the human mind. As that becomes more developed, more enlightened, as new discoveries are made, new truths disclosed, and manners and opinions change with the change of circumstances, institutions must advance also, and keep pace with the times. We might as well require a man to wear still the coat which fitted him when a boy, as civilized society to remain ever under the regimen of their ancestors.

—Thomas Jefferson

LIMITS OF THE EQUILIBRIUM MODEL IN SOCIAL RESEARCH

by David Easton

Department of Political Science, University of Chicago

An analysis is offered of the utility of an equilibrium model for the study of social systems. In some areas of social science research is conducted explicitly in terms of an equilibrium model, as for example, in economics. In most disciplines, however, it lurks beneath the surface as an inarticulate theoretical framework, as is the case in political science. An attempt is made to clarify the meaning usually associated with the concept equilibrium and to distinguish among various types of equilibria. The major shortcoming of this model at the present stage of development in social research derives from the lack of data quantifiable on either an ordinal or a cardinal scale. Only to the extent that such measurement is available can equilibrium analysis be fruitfully used to understand the operations of an empirical social system. Otherwise the model remains formal and untestable.

THE purpose of this paper is to inquire into the extent to which the equilibrium model is a useful mode of analysis for understanding social behavior.[1] As a central concept, the idea of equilibrium recurs persistently in the various social sciences and it represents perhaps one of the few analytical orientations common to all social research. In economics, for example, it lies at the heart of the discipline, providing the basis for a general theory of economic life. In none of the other social sciences, of course, does it play so central a role or has it been elaborated in so sophisticated a way. Nevertheless, even though the concept has not been consciously adopted as a pivot for analysis, implicitly at least it has found its way into research. Indeed it has become so deeply imbedded in on-going research that it is seldom extricated for sustained inquiry into its appropriateness or utility.

In sociology and anthropology, for example, many texts and even advanced theoretical monographs make use of the idea of equilibrium in some important way (2, 5, 7). And in political science, although the concept usually appears under the guise of such synonyms as adjustment, accommodation, or the balance of power, in fact the idea of equilibrium stands as the closest approximation to a general theory that can be found in the whole field.[2] And what is even more to the point, sustained pressure continues from the direction of biology and psychology openly advocating or implicitly urging that the biological, individual, and social systems might best be interpreted in some kind of equilibrium term.[3]

The equilibrium model

The argument favoring this position runs in somewhat the following way. All scientific research, it is maintained—and with considerable justification—ultimately deals with determinate systems of behavior. From this premise, the further conclusion is developed that since all systems are determinate, they must show the property of striving to achieve equilibrium, variously called steady state, homeostasis, adjustment, balance, and the

[1] I am indebted to Professor Jacob Marschak, until recently a colleague at the University of Chicago, for his kind advice on a number of points. He is of course blameless for my general orientation or specific conclusions.

[2] See Reference 6, especially Chapter 11.

[3] See Reference 14.

like. For this reason it seems logical to the advocates of this position to propose first, that scientists interested in developing the most general model or theory for understanding all kinds of human behavior ought to devote themselves to a study of the properties common to all systems—general system theory as it has been called; and second, that this inquiry into system theory must be couched in the form of equilibrium analysis (1, 3, 4, 10).

The prevalence or the growing popularity of the equilibrium idea, however, in itself offers little evidence of its necessary utility either for revealing reliable knowledge about a specific area of behavior to which it may be applied or as a general theory of social behavior. As a matter for independent discussion it is worth posing the following question: To what extent can equilibrium analysis be profitably employed at the present level of development in the social sciences? In general my conclusions will be, first, that for purposes of scientific research, it is useful to approach societies and their major segments as systems of social behavior, as suggested by the equilibrium mode of analysis; second, that even though we may find this rewarding, systems theory does not require premature commitment to the equilibrium approach; and third, that equilibrium analysis is suitable only for throwing light on some limited aspects of behavior, at least at the present stage of technical development in the social sciences. At present it has inescapable shortcomings as a general theory of behavior or for that matter as a more limited theory to guide research in any one field such as politics, culture, or social structure.

Before engaging in further discussion I would like to recall one postulate which I believe is implicit in most scientific research. This is that all those variables essential to an understanding of a significant area of social behavior constitute a determinate system. Following W. R. Ashby in *Design for a Brain*, I shall call the interacting elements within a system, the variables and those elements outside the system or in the environment, the parameters. On this postulate and definition I shall now inquire into what we must mean when we talk about a social system being in equilibrium.

In the literature of social research, any sampling of meaning associated with the concept of equilibrium would display a considerable degree of obscurity and conceptual indecision, so much so in fact that there would be little point in itemizing the various shades of meaning loosely attached to the term.[4] Such an inquiry would, however, yield at least one negative conclusion. Normally the social scientist does not intend to equate the equilibrium in a social system with one in thermodynamics. In thermodynamics, according to the familiar second law, every system is assumed to have a fixed quantity of energy. In the process of interaction among the elements, energy differences tend to disappear until the system approaches what has been called entropy death. All forms of energy are transformed into heat and heat differentials are reduced to a common level. In describing a social system as a state of equilibrium, social scientists usually do not intend to suggest that the system is characterized by entropy. Implicitly at least, social scientists act on what I would consider to be an indisputable premise that there is a large store of energy available to any social system and accordingly entropy cannot occur. And even if it could be demonstrated that systems characterized by entropy do occur, this would not shift the main burden of my remarks. We need simply consider these systems as special cases of equilibrium. As such, whatever limitations are found in the general utility of equilibrium analysis would apply with equal force to the thermodynamic case if and when it appears in social systems.

Even granting this negative agreement concerning the use of the idea of equilibrium, that is, that it does not imply the occurrence of entropy, it is nevertheless impossible to discover any definition that will in fact embrace the appalling varieties of meaning with which the term appears in social research. But if I can generalize from social science

[4] The almost impenetrable confusion and variety of meaning has already been exhaustively explored. See Reference 13, Volume IV, Chapter 14.

usage, without too great risk, I would say that in its most general sense the term equilibrium has real significance only if it is understood to imply the following: In every system the component variables will interact in such a way that if the interaction is allowed to continue without further disturbance, ultimately a state will be achieved in which no variable changes its position or relation with respect to the other variables. In this sense only is it meaningful to say that variables are in equilibrium, that they have adjusted to each other, that they have reached a steady or homeostatic state, or that they enjoy a condition of harmony, stability, or balance.

The social equilibrium is different from the thermodynamic primarily because where human beings are involved activity does not cease, even under conditions of equilibrium. Nor for that matter, although this is not usually clearly recognized, need change cease. All that the state of equilibrium demands is that the pattern of interaction at the center of attention becomes fixed. If in a given interval of time the relationships among the variables remain fixed, we can say that the system is in equilibrium with regard to these variables.

We can appreciate more fully the meaning of a state of equilibrium and why it does not necessarily involve the cessation of activity or change if we look for a moment at the two broadest categories of equilibrium. Conventionally equilibria are classified as stationary or dynamic. A stationary equilibrium occurs when there is equilibrium and the parameters of the system are held constant over a given interval of time. For example, all the groups in a political system compete for power over the establishment and execution of political policy. If and when ultimately a policy is established in which all groups acquiesce, the political system has often been described as being in balance or equilibrium. We can describe this state as a stationary equilibrium if we assume that the environment is held constant and therefore that factors that influence the size, composition, and general power of political groups, such as technology, preferences, social structure, and size of population, do not change. If the power of each competing group is constant, even though the group might like to alter policy, nevertheless under the circumstances each group feels it has gotten the most it can and that it cannot improve its position. Hence, no group moves out of its position and a point of balance or equilibrium is achieved. Activity, however, continues in the sense that each group continues to exert its available pressure.

The same holds true for dynamic equilibrium, the second major kind of equilibrium. Here, of course, not only does activity persist, but change takes place as well. For this reason the concept of dynamic equilibrium is adapted to the study of historical change in social systems. What makes the historical problem different is the fact that the parameters (or environment) of the system under discussion undergo constant change. The result is that the position of equilibrium of the system itself is always shifting. However, if the rate of change is itself constant, or if the change has a pattern that consistently repeats or maintains itself, then the process of change is said to have achieved a state of equilibrium. If, for example, it could be demonstrated that civilizations rise and fall in a fixed repetitive pattern over the ages, then we could say that the life spans of civilizations had arrived at a state of dynamic or moving equilibrium. This state is one that does not change if by this we mean only that the pattern or rate of movement is fixed and repetitive over a given interval of time. Equilibrium, therefore, whether viewed as a stationary or dynamic state, need not bring a halt either to activity or change respectively.

There is one meaning so frequently attributed to equilibrium that I must say a special word about it. It is quite common to find equilibrium described as a condition in which, if the environment is held constant, and if any variable is then disturbed and released, the system will display a tendency to return to the initial state.[5] This is a substantially different behavior from my earlier description of equilibrium as a state in which the relationships among the participating elements remain constant. In fact, if it can

[5] See, for example, Reference 7, pp. 303–304.

be shown that a given system tends to return to its original state when it is displaced, what we have is a special kind of equilibrium. Usually this is called a case of stable equilibrium as contrasted to an unstable or neutral equilibrium. To describe equilibrium in general by the special properties of a stable equilibrium would result in unnecessarily confining all equilibria to this one case. We would be substituting an instance for the whole class.

Varieties of equilibria

To clarify further the difference between equilibrium in its generic sense and its meaning in the special case of stability, we shall have to look for a moment at the possible varieties of equilibria. For purposes of brevity and clarity, however, I shall confine my illustrations to the stationary equilibrium of physical rather than social phenomena.

A stable equilibrium is illustrated by a ball at the lowest point of a spherical cup. If it is displaced within given limits from its position of rest in the cup and released, it will return to its initial position. However, the ball would be in a condition of unstable equilibrium if it were poised on the hill joining the sides of two adjacent cups. Any slight disturbance would drive the ball from its initial position down into one or the other of the cups where it would finally rest. The original position was unstable because a slight disturbance evoked further disturbances as the ball gained momentum down the slope and this continued in a cumulative way until the system took up a new position of equilibrium. Finally, a ball would be in neutral equilibrium if its initial state were one of rest on a flat surface with sufficient friction to stop the ball once it was set in motion. Then a disturbance would simply move the ball to a new position of equilibrium without setting in motion any snowballing effects. The original equilibrium was neutral because any disturbance immediately resulted in the ball adopting a new position of rest.

The reason why in social research equilibrium in general is often mistaken for and described in terms of the properties of a stable equilibrium is perhaps because most research workers agree, rather instinctively, that social systems are stable. However, it is important to bear in mind that stability in this sense would be a characteristic of only one kind of equilibrium. And in fact there is no a priori reason why any or all large social systems if in equilibrium at all, far from showing properties of stability, might not be either unstable or neutral equilibria. In the light of the paucity of serious equilibrium research with regard to social systems, it is certainly premature to characterize all such systems as stable.

In exploring the meaning of equilibrium, there is perhaps one further type of equilibrium to which some attention ought to be devoted if only because the biologist Bertalanffy has made a strong case for utilizing it as the central idea of general systems theory. Bertalanffy designates the equilibrium found in the human organism as a steady state. The peculiarity of this kind of equilibrium lies in the fact that its point of rest is independent of the path taken by the variables. Thus, in order to survive, the human organism must maintain certain essential functions within specifiable critical limits. Blood pressure, sugar content of blood, body temperature, and other variables must be held within a limited range. Disturbances of these states may arise from a number of directions and what distinguishes the human organic system from, say, the thermodynamic system, is that the same end, namely, a steady state, can be achieved from a variety of different initial states, regardless of the path followed. Bertalanffy calls a system with this property equifinal. The final result is the same whatever the mechanisms or paths required to achieve it.

What is often neglected in connection with the equifinal steady state is that this is a peculiarity only of one kind of system and is not necessarily true of all systems, such as the social. Indeed, economists have long been able to demonstrate that the path taken towards equilibrium may in fact influence the final position of rest; at times it may even forever prevent the attainment of equilibrium. Perhaps it is true that all organic systems are equifinal but we already have

evidence from economic theory that this is not necessarily the case under all circumstances for social systems. Only in certain instances in economics is the final position independent of the route followed (8).

It is clear, therefore, that if we are to be exhaustive about what we mean by equilibrium we must avoid applying the description of a special case to the whole field, at least in the social sciences. Neither the description of a stable equilibrium nor that of an equifinal steady state serves to represent the most general meaning of the term. For our purposes, whether any instance in social interaction is an example of a stationary or dynamic equilibrium, whether it is stable, neutral or unstable, whether it shows the property of a thermodynamic system or of an equifinal steady state, we shall consider the human participants to be in equilibrium when no individual or group changes its position vis-à-vis the other participants. This meaning of equilibrium accords with its usage today in physics and economics at least. Whatever I have to say about the equilibrium of a system will apply to the concept in this, its most inclusive sense.

Value of the equilibrium model

Having clarified the general meaning of equilibrium, we are now ready to ask: What is the essential value of equilibrium analysis? It does not of course lie only in the fact that this form of analysis encourages us to interpret significant areas of behavior as systems or determinately rather than randomly related sets of variables. The idea that for purposes of analytic research important segments of behavior can be viewed as a system, is not unique to the equilibrium approach. Other types of models are founded on the same premise. The central significance of the equilibrium concept derives rather from the fact that it offers a rewarding, but for all that only *one* rewarding way of understanding the mutual relations among the elements of a system.

The core of equilibrium analysis and the source of its utility for helping us to understand social behavior lies in the kind of leading question it prompts us to ask. To employ the notion of equilibrium we must initially seek an answer to the following question: What are the conditions of equilibrium? Or alternatively, what are the multiple relations among the variables if the system under study is to achieve equilibrium? These relations are designated as the laws of the system. For example, if price, supply, and demand are identified as three of the essential variables of the economic system, equilibrium research requires that their relationship be clarified. The so-called laws of supply and demand specify one of the ways in which these variables interact.

Once we know what the conditions of equilibrium are, we hold in our possession a set of generalizations from which a vast body of deductions may flow. This deductive quality of equilibrium analysis is its ultimate justification as a theoretical model. Knowing the relations of the variables in a state of equilibrium, it becomes possible, for example, to explore formally the way in which a disturbance of one of the variables influences the behavior of the system: What effect has it on the state of equilibrium? Does the system return unimpaired to its initial state or is the subsequent equilibrium position changed? How fast does the system move to its new state? To what extent does the position of equilibrium depend upon the route followed in the period of adjustment after the displacement? In short, the idea of an equilibrium, like any other mode of analysis, simply helps us to unearth invariant relations among the components of a system. Its value stands or falls on this ability.

In spelling out the value of equilibrium analysis in this way, I have assumed for the moment that any given social system must in fact reach a point of equilibrium. There is, however, no valid ground for making this assumption; whether or not a social system does achieve equilibrium must always be a matter for empirical investigation. Nevertheless, even if we should assume that empirical systems never do arrive at an equilibrium, that, for example, the parameters of any significant social system are in constant and irregular flux, this would not necessarily destroy the possible utility of this kind of analysis. It would still be valuable to help us search for the invariances that abort equi-

librium in a system. The objective of equilibrium research in this case would be to specify the speculative conditions of equilibrium, that is, how the variables of the system would have to be related in order for equilibrium to occur. We would then have a norm against which to check and compare the actual systems. The theoretical state of equilibrium would then become a point of reference to help us interpret the real world by understanding the actual divergences from the norm.

For example, it is sometimes maintained that the economy is never in fact in equilibrium but is rather in constant disequilibrium, moving towards and away from equilibrium or always in the "neighborhood of equilibrium."[6] The concept of equilibrium here becomes purely a heuristic device designed to simplify research into the complex economic process. By spelling out the laws governing the interaction of economic variables if they were in a state of equilibrium, it becomes possible to compare the real economic system against the theoretical norm. Any divergence between the two is then explained as a result of the operation of certain factors which theory tells us would account for the distance of the real system from a point of equilibrium.

Limitations of the equilibrium model

Whether the equilibrium model is conceived as an actual description of a social system or only as a theoretical norm, no one could possibly deny the manifest potential assets flowing from its use, as I have just indicated. The only question I would raise here is whether in the social sciences, at their present level of development, it is possible to realize upon these potentialities. Now, in order to answer this doubt we must be prepared to recognize not only the possibilities of this model but also its limitations. And when we come to look at these limitations we immediately see that if equilibrium concepts were adopted in the social sciences as a primary form of analysis, except for special cases to be specified later, it might easily trap current research in an unsuspected and dangerous blind alley.

The reason for this can be simply stated. By its very nature the application of an equilibrium model to actual social situations requires careful quantification of the relevant variables. Where this is possible, as for example in economics, or in the case of small face-to-face social groups, the equilibrium model will yield its manifest dividends. It may be successfully employed in these cases, however, only because there the central magnitudes, such as supply and demand in economics, are measurable. Where, however, quantification in any of the senses indicated in the next paragraph is as yet and in the foreseeable future, impossible, it would be a waste of valuable time, energy, and other resources to confine ourselves to this mode of analysis.

The idea of quantification is used here in its broad sense. It includes all types of ordinal as well as cardinal measurement. For purposes of equilibrium analysis, measurement on an absolute or cardinal scale is of course unnecessary; ordinal scaling can serve equally well. For example, if we can firmly establish the rank order of a set of variables and if after a disturbance the same rank order appears, we can then say that the system is in equilibrium, at least with regard to the rank ordering.

The reason why measurement so defined is needed for purposes of equilibrium analysis is that without such measurement we would find it impossible to tell when the empirical system is in fact at a position of equilibrium. Furthermore, we would not be able to answer other significant questions raised by the equilibrium model, such as these: how far is the system from equilibrium, how long does it take to reach this state, and so on with regard to an almost unlimited list of quantitized questions that give equilibrium analysis its significance. In short, we would lack the important numerical values with which to fill in the unknowns in the mathematical equations which would constitute the statement of relations in the formal analysis. Most of the important deductions in which the equilibrium model is rich, would

[6] For a concrete illustration of this danger, with regard to political science, see Reference 11, especially Chapter 2.

be unverifiable by reference to the empirical system.

This shortcoming can be easily illustrated. Suppose we were studying a primitive society such as the Trobriand Islands. How would we know when such a society, conceived as a social system, was in stable equilibrium; that is, that after a disturbance none of the essential elements changed its position. Impressionistically we might of course determine that the relations among the Trobriand Islanders were of such a nature that if any external stimulus were brought to bear upon the system and then withdrawn, the initial relations would immediately tend to reassert themselves. That is, we might be able to determine through observation whether the Trobriand society seemed to have the properties of a stable equilibrium. To this point, our observations and inductions, like all initial insight, would be satisfactory even though cast in unquantified qualitative terms. But beyond this point we would become caught up in the limitations of an unquantified use of equilibrium analysis.

Although initial observation might tell us that the Trobriand society was in a state of equilibrium, we would in fact be unable to demonstrate this beyond reasonable doubt. It might *appear* that the system returned to its initial state after the disturbance was withdrawn. But unless we could fortify this insight with either cardinal or ordinal measurement of each variable entering into the relations necessary to evoke equilibrium, we could not determine whether these variables in fact returned to their initial state of interrelationship. We could say only that they appear to do so. Since we know how deceptive appearances are in scientific research, we could easily be misled by our impressionistic observations. It is quite conceivable that the relations among the variables had changed but that these changes were imperceptible without measurement. Unless we could specify precisely how much of each variable must be related to how much of each other variable for the system to be in equilibrium or what the rank order of the variables is, we would never have reliable knowledge about the conditions under which this state does in fact occur. And until we knew this, most of the manifest dividends of equilibrium analysis would remain beyond our reach.[7]

Let us take another example. Suppose that the tariff policy of the United States at a given time is characterized as the equilibrium policy established through the interaction of a given number of pressure groups. By this we would mean that with regard to the power being exerted on tariff policy these pressure groups are, let us say, in stable equilibrium. As an insight this formulation appears very fruitful. But when we come down to verifying the insight, that is, to applying the theory, unless we have some cardinal or ordinal measure of power it would be impossible to confirm that any disturbance led to the reassertion of the pre-existing power relations.

If quantification were possible, we could demonstrate the re-establishment of these relations in at least two ways. First, if we had a measure of power based upon an absolute or cardinal scale, we would of course find little difficulty in determining whether each group continued to exert the same amount of power over tariff policy. Second, even if we lacked cardinal measurement, we could still speak about the return of the pressure groups to their original relationships. Such a statement would be possible if we had established a rank order among the various pressure groups indicating their relative positions in terms of power. Then if after the disturbance the same rank order appeared, we would be correct in stating that the system of variables we call "the power of pressure groups in relation to tariff policy," was in stable equilibrium.

But the fact is that political scientists have as yet developed no reliable measure for the amount of power held by pressure groups; in this shortcoming lies the true barrier to analysis of the power relations of these groups in equilibrium terms. It is quite apparent that there is no satisfactory unit or index for discovering the exact amount of

[7] Since writing this article a fascinating study has appeared that vigorously challenges the utility of empirical research guided by the model of a stable equilibrium. See Reference 9.

power, in cardinal terms, held by any group or individual. But what is even more important and less well recognized is that, at the present level of knowledge in the social sciences, it is impossible even to rank pressure groups on some relative power scale. In some limited cases, as, for example, where the differences in power are manifestly great, an acceptable intuitive subjective ranking may be possible and satisfactory. But where the social groups under consideration may appear to lie close together on an ordinal power scale, subjective ranking becomes too unreliable. Until we know more adequately the kind of phenomena to which we are referring when we speak of power, and until we have developed indices at least for ranking various degrees of power, we must find it impossible to speak even about an equilibrium of rank-order relationships.

It is precisely such conditions that we find most frequently in the social sciences today. Our terms are not sufficiently refined or differentiated to enable us to speak with assurance about the quantitative aspects of the variables to which these terms refer. Under these circumstances, I am arguing, and to the extent to which they persist, equilibrium analysis has doubtful value. It must remain at the level of insight rather than carefully elaborated and verifiable theory.

To be perfectly clear about what this conclusion means, two points must be made. First, the fact that the variables of most social systems are not as yet quantifiable does not prevent the construction of a purely formal or mathematical equilibrium model. Measurement becomes necessary only when we seek to apply the formal model to reality. At the theoretical or formal level, a model for any social system might conceivably be developed through some form of mathematical representation, emulating perhaps the economic model of perfect competition and its various modifications. Here since we would not initially be concerned with applying the theory to concrete social behavior, we could ignore the lack of quantifiable major variables. We could trace out mathematically what would happen under a variety of hypothetical conditions without worrying in the least whether the concrete variables into which our symbols must eventually be translated were in fact measurable. This is the prime virtue and at the same time the inherent danger lurking within the mathematical treatment of systems; as someone has said, it enables us to be perfectly clear and precise without however being specific. The concealed danger is that when we do seek to become specific, when we test our analytic framework by applying it to the empirical system, if we find the variables lack numerical values, our theoretical work will have been, if not in vain, at least grossly premature.

The second point that must be noted is that this fundamental shortcoming of equilibrium analysis does not put it beyond the pale for all social research. Many of its unique merits as a model *are* accessible. This is true, of course, where because of the nature of the social system involved, the variables can be quantitized. But the advantages of equilibrium research are also at least partly available when the existence of equilibrium can be ascertained by inspection. For example, even if we could not at first measure the elements of some systems, it might still be possible to establish the position of equilibrium. This is true of certain physical systems such as a ball at rest in a spherical cup. We can test whether it is in equilibrium through simple direct manipulations and observations that require no buttressing by measurement, at least when we approach this system macrophysically.

Similarly with regard to some social systems, mere inspection can inform us of the stability of a set of relations. This is possible, for example, in the case of such small face-to-face groups as those studied in the research work conducted at the Western Electric Company's Hawthorne Works in Chicago from 1927 to 1932. In small groups the relationships are directly visible at one time and can be manipulated without too great difficulty. Even without measurement something can be learned from such small groups about the conditions of equilibrium. However, it often happens that just because these social systems are small and manipulable, we are able to devise techniques for measuring the important variables. For example, the

conditions of equilibrium of small groups can be explored through the measurement of the frequency, scope, and rate of interaction among its members (12). To the extent that measurement is at present possible or within reach at some none too distant future, the equilibrium model offers a valuable analytical tool.

If however we are concerned with developing a general theory for the study of human behavior or for the understanding of any large segment of the social system, such as its structure, culture, or politics, then the requirement of quantifiable variables as I have defined measurement, acts as a serious, and yet at this level of development in social research, as an inescapable barrier to the use of an equilibrium model. The model could be developed, as I have already indicated, in mathematical or purely formal terms, but it would fail to meet one of the criteria of a good theory, namely: it would not be readily applicable to concrete empirical situations.

This conclusion does not, of course, relegate equilibrium analysis permanently to the limbo of social research. As time goes on and as our techniques and insights develop, there can be no doubt that an increasing number of social variables will yield to precise identification and measurement. At the moment, however, our inability to measure some of the most significant dimensions of social behavior suggests that if we are seeking to understand the relations among such dimensions, we would be more richly rewarded if we devoted more of our resources in directions other than that of equilibrium analysis. Careful research requires that in advance we clarify the areas in which equilibrium analysis may possibly be most fruitful. And where equilibrium analysis is less rewarding at the present level of concepts and techniques, we must be willing and content to experiment with other models that can be applied without excessive concern for immediate quantification, whether cardinal or ordinal.

In conclusion then, although we may seldom realize it, the application, of equilibrium analysis to broad aspects of the social system is normally a subtly dangerous practice. It leaves the impression that we have a useful general theory when in fact, lacking measurability, it is a mere pretence for knowledge. It leaves the impression that we are saying something significant about the social system, when in fact careful scrutiny reveals that equilibrium terminology often conceals the absence of genuinely usable theory.[6]

REFERENCES

1. Ashby, W. R. *Design for a brain*. New York: Wiley, 1952.
2. Bennett, J. W., & Tumin, M. M. *Social life*. New York: Knopf, 1949.
3. Bertalanffy, L. v. An outline of general system theory. *Brit. J. Phil. Sci.* 1950, 1, 139–164.
4. Bertalanffy, L. v. *Problems of life*. New York: Wiley, 1952.
5. Chapple, E. D., & Coon, C. S. *Principles of anthropology*. New York: Henry Holt, 1942.
6. Easton, D. *The political system*. New York: Knopf, 1953.
7. Homans, G. C. *The human group*. New York: Harcourt, Brace, 1950.
8. Kaldor, N. Note on the determinateness of equilibrium. *Rev. econ. Stud.*, 1934, 1, 122–136.
9. Leach, E. R. *Political systems of Highland Burma*. London: G. Bell and Sons, 1954.
10. Pareto, V. *Mind and society*. New York: Harcourt, Brace, 1950.
11. Schumpeter, J. A. *Business cycles*. New York: McGraw-Hill, 1939.
12. Simon, H. A. A formal theory of interaction in social groups. *Amer. sociol. Rev.*, 1952, 17, 202–211.
13. Sorokin, P. *Social and cultural dynamics*. Chicago: American Book Co., 1941.
14. Weber, C. O. Homeostasis and servo-mechanisms for what. *Psychol. Rev.*, 1949, 56, 234–239.

Our earth is but a small star in the great universe. Yet of it we can make, if we choose, a planet unvexed by war, untroubled by hunger or fear, undivided by senseless distinctions of race, color or theory.

—Stephen Vincent Benét

ADAPTIVE BEHAVIOR FROM RANDOM RESPONSE

by Donald T. Campbell

Department of Psychology, Northwestern University[1]

Darwin has shown how the seemingly purposeful process of evolution could be explained by the piling up of random variations properly selected. Applied to learning this principle bridges the gap between the ideas based on mechanical principles of behavior and those based on "intelligence" or "purposefulness," much the same way as this gap is being bridged in servo-technology. It is suggested that all existing learning theories contain explicit or implicit assumptions about some selective principle operating on initially random responses, assumptions which Ashby has carefully spelled out and utilized in the construction of his Homeostat.

CAN *habit* be effectively treated as a semi-isolated stable system, à la Ashby (1), or as a low entropy high energy steady state, à la Bertalanffy (2), or as a homeostatic system, after Cannon (5)? Such questions are timely, in that the influence of cybernetics, servo-mechanism models, and general systems theory are suggesting such reformulations to a number of psychologists. But such timeliness is not sufficient to justify the enterprise. In seeking new models in psychology we should be motivated by something other than the desire to pour old wine into new jugs. And in particular when we come to deal with the concept of *habit* special justification seems necessary, inasmuch as we are dealing with the oldest, most respectable, and most adequately handled of psychological concepts. What are some of the difficulties with our present conceptualization of *habit* which tempt us into this new family of models?

Most important for the present discussion has been the difficulty in the definition of the character of the acquired response. Two principal operational alternatives have offered themselves: One of these is to define the response in terms of the body's own coordinates, in terms of muscle contractions or movements. The second alternative, equally operationally specifiable, has been to define the learned response in terms of a change in the relationship between organism and environment. Responses defined in these terms become objects moved, places reached, regions entered, etc. It is characteristic of the great bulk of research on habit that these two alternative definitions of the learned response have been confounded. That is, no operations have been provided in the experimental situation which would distinguish between these two interpretations, since learning is continually tested in the identical situation in which the habit was acquired. This has given learning theorists the possibility of overlooking the existence of the alternatives, and of arbitrarily assuming one of the definitions in the process of building their theories. In the few experimental studies in which these two interpretations have been operationally opposed, in which through transposition experiments two interpretations have been unconfounded, the evidence is predominantly but not entirely in favor of the second interpretation (4, 18). That is, the learned response is to be essentially defined in terms of a shift in the organism-environment relationship rather than a motor response defined in terms of organism or body parameters alone. The most popular and best developed learning theories have ignored this fact and have gone ahead assuming the adequacy of

[1] This paper was made possible through the author's participation in the Committee on Behavioral Science, University of Chicago, 1952–53.

defining response in isolated motor or muscle terms, even when using molar data collection units.

A major reason for this resistance is that accepting the validity of the second definition of response presents the learning theorist with a much more difficult problem, a problem which cannot readily be encompassed by the anatomically reinforced conceptualization of the simple reflex arc, with discrete afferent and efferent neural transmission pathways. A second reason for resistance lies in the fact that those who, like Tolman (23) and Lashley (14), have pointed out the adaptive, object-consistent nature of learned responses, have not presented plausible mechanistic theories of how such responses could be achieved, and have used language implying purpose, thus raising the ghost of metaphysical issues which had been important at the time when psychology was freeing itself from philosophy. Recent attention to the engineers' achievement of self-regulating machines, the behavior of which is consistent to uniform ends achieved through variable intermediate steps (as in the thermostat and governor at the most simple level and as in the aiming and tracking devices for anti-aircraft artillery at a more complicated level), has supplied a mechanistic model of adaptive, habitual response mechanisms. While we cannot say that adopting a steady state or homeostatic model will solve at the psychological or physiological level the problem of how such responses are achieved by the organism, it does at least give a deterministic model for facing squarely one of the toughest aspects of the actual phenomena of habit.

In a recent book, Ashby (1) has proposed the minimum elements for a machine which would show adaptive behavior. As indicated in his title *Design for a Brain* it is his explicit intent to develop such a model for the furtherance of the understanding of the adaptive behavior of living organisms. While it is not the purpose of this paper to evaluate the over-all success of his achievement or its total relevance for psychology, the assignment which he gives himself is an important one, and directly pertinent to our problem.

> How does the brain produce adaptive behaviour? In attempting to answer the question, scientists have discovered two sets of facts and have had some difficulty in reconciling them. The physiologists have shown in a variety of ways how closely the brain resembles a machine: in its dependence upon chemical reactions, in its dependence on the integrity of anatomical paths, and in many other ways. At the same time psychologists have established beyond a doubt that the living organism, whether human or lower, can produce behaviour of a type called 'purposeful' or 'intelligent' or 'adaptive'; for though these words are difficult to define with precision, no one doubts that they refer to a real characteristic of behaviour. These two characteristics of the brain's behaviour have proved difficult to reconcile and some workers have gone so far as to declare them incompatible.
>
> Such a point of view will not be taken here. I hope to show that a system can be both mechanistic in nature and yet produce behaviour that is adaptive. I hope to show that the essential difference between the brain and any machine yet made is that the brain makes extensive use of a principle hitherto little used in machines. I hope to show that by the use of this principle a machine's behaviour may be made as adaptive as we please, and that the principle may be capable of explaining even the adaptiveness of Man (1, p. 1).

As a first part of his enterprise, Ashby has built a machine which he calls the "homeostat." While this machine is but a fragment of his total model it nonetheless manifests certain adaptive or problem-solving characteristics, constituting what Ashby calls an ultrastable system. It is one feature of this system, repeated in his full model, to which attention is turned in the present paper.

> When these parameters are given a definite set of values, the magnets show some definite pattern of behaviour; for the parameters determine the field and thus the lines of behaviour. If the field is stable, the four magnets move to the central position, where they actively resist any attempt to displace them. If displaced, a co-ordinated activity brings them back to the centre. Other parameter-settings may, however, give instability; in which case a 'runaway' occurs and the magnets diverge from the central position with increasing velocity . . . when, and only when, the magnet *M* diverges far from the central position . . . the coil *G* of each uniselector is energised . . . ; for only at extreme divergence does the output-current reach a value sufficient to energise the relay *F* which closes the coil-circuit. A separate device . . . interrupts the coil-circuit regularly, making the uniselector move from position to position as long as *F* is energised. . . . The uni-

selector (or 'stepping-switch') U . . . [controls the parameter settings]. . . . The values of the components in U were deliberately randomised by taking the actual numerical values from Fisher and Yates' Table of Random Numbers. Once built on to the uniselectors, the values of these parameters are determined at any moment by the positions of the uniselectors. Twenty-five positions on each of four uniselectors (one to each unit) provide 390,625 combinations of parameter-values. [Thus eventually a setting is achieved which returns the magnet to its central position, inactivating relay F, and the uniselector is therefore left at that setting.] (1, p. 96).

It is not anticipated that this brief sample from Ashby will give the reader an understanding of his machine and the requirements he sets for it. It is doubtful that such understanding can be provided more economically than Ashby himself has done. But the quotation may be adequate to convey to the reader the essential gimmick in the machine which makes it adaptive. When the machine is under stress, this stress sets into operation a mechanism which emits randomly various trial settings or behaviors until a setting occurs which removes the stress and thereby interrupts the emitting of further settings.

To one who has picked up Ashby because of his promise to deal with "purposeful" or "intelligent" or "adaptive" behavior, encountering this arbitrary and random little gadget may come as something of a letdown. One might be tempted to apply Kurt Lewin's criticism of G. E. Muller's associationism:

> The explanation, following closely Darwinistic ideas of directed actions as the result of factors which do not contain the concept of directedness, was considered a particularly important achievement because at the time the controversy between the teleological and causal explanations of behavior was acute. The outstanding characteristic of a scientific causal explanation was incorrectly seen as requiring the avoidance of the concept of direction. This view was held in spite of the fact that one of the basic concepts in physics, the concept of physical force, refers to a directed entity (a vector, in terms of mathematics) (16, p. 66).

But because Lewin may have underestimated the confusion remaining in psychology on issues of teleology and conscious purpose, and because the Darwinian conceptual achievement in dealing with adaptation on the phylum and species level is not one to be despised, we may be justified giving Ashby further consideration. He himself recognizes the Darwinian roots of his model:

> The work also in a sense develops a theory of the 'natural selection' of behaviour-pattern. Just as, in the species, the truism that the dead cannot breed implies that there is a fundamental tendency for the successful to replace the unsuccessful, so in the nervous system does the truism that the unstable tends to destroy itself imply that there is a fundamental tendency for the stable to replace the unstable. Just as the gene-pattern, in its encounters with the environment, tends towards ever better adaptation of the inherited form and function, so does a system of step- and part-functions tend toward ever better adaptation of learned behaviour (1, p. vi).

If indeed the engineer attacking the problem Ashby has set himself has no alternative solution, and Ashby would claim that this is essentially so, then the device in question and its psychological parallels seem worthy of being taken quite seriously by the psychologist. Furthermore, it turns out upon inspection that most major psychological theories embody such a mechanism, and that Ashby's approach, if nothing else, may serve to point out one major point of agreement among our competing schools of learning. To make this assertion it is necessary to modify somewhat the demand for perfect randomness, and indeed, this is not essential to Ashby's argument. The essential mechanism becomes one of continual variation of behavior in the face of continued unresolved motivation. Randomness, or better, blindness, remains in that although the responses may have biased initial likelihoods of being used, the successive efforts or trials do *not* have the character of "corrections" of previous trials. Thus while eventually a more accurate or adaptive response is achieved by chance, there is no relationship between the direction of miss on one trial and the character of the next.

Thorndike (20) and Hull (11) enter their organism into the situation with an initial

family of divergent responses associated to the situation in varying strength.

$$S_1 \begin{cases} \underline{p. = .05} \; R_a \\ \underline{p. = .45} \; R_b \\ \underline{p. = .15} \; R_c \\ \underline{p. = .20} \; R_d \end{cases}$$

If R_b does not result in reinforcement, its likelihood of occurrence is successively diminished until the next strongest response, R_d, occurs, and so on until a reinforced response (possibly, though not necessarily, one which eliminates S_1,) occurs. The overt behavior described is compatible with that of Ashby. In the limited sense described, the responses vary at random, or blindly. The conditions of termination of response alternation are essentially the same. The differences are worth noting too. There is no S_1 term in Ashby's model, and no unequal prior response likelihoods—although the latter could get built into the uniselector without changing the essential mode of operation. Furthermore, the retroactive effect of reinforcement in strengthening the bond is not used by Ashby.

Skinner's theory (19) of learning is still more similar. His concept of emitted responses, bound by no prior association to specific external stimuli, is quite reminiscent of Ashby's gadget. However, Skinner's use of reinforcement as increasing reflex reserve departs along lines similar to those of Hull and Thorndike.

Guthrie's (7, 8, 9) description of puzzle box behavior most closely parallels Ashby's machine—both as to the random emitting of responses, and as to the mode of selection of the adaptive response. For Guthrie, the response that removes the animal from the stimulus situation that instigated the random floundering is strengthened because it is allowed to remain unoverlayed with the traces of other responses.

It is harder to point to random response-emitting mechanisms in those theories influenced by Gestalt psychology. In general, where response variability in the face of a continued unsolved problem is discovered, there is an emphasis upon the adaptive, intelligent, corrective nature of these responses, not their random character. (This distinction is made clear in Hilgard's contrasting "trial and error" concepts with those of the "provisional try" and "approximation and correction" (10, pp. 335–339).) However, it is felt that where Gestalt theories conscientiously cover learning in entirely novel situations, random behavior enters the description at some point. In 1925 and 1926 Tolman (21, 22, 24), in partial agreement with earlier statements by Perry and McDougall, was describing purposive behavior in terms of the same essentials that Ashby is now using. In Tolman's terms the essential character of such behavior is "persistent . . . random trial and error . . . *until*" (22, 24, p. 50). While the randomness of the trial and error has not been a central emphasis in Tolman's over-all theory, it is nonetheless clearly present in a form which Ashby has closely paralleled (21, 22, 23 pp. 339–341). N.R.F. Maier (17) specifies a natural random variability as precursor to learning. The "discontinuity" theories of Krechevsky (13) and Lashley (15) imply a succession of problem solution efforts which are independent of each other, and hence do not manifest the cumulative character of successive corrections. Discontinuity would seem to imply randomness or independence between successive "hypotheses." Certainly in the original "hypotheses in rats" study (12), where blind alleys were assigned at random on successive trials, the hypotheses that were emitted by the rats were randomly or blindly emitted, in the sense of being independent of previous hypotheses: in the sense of not being "corrections" of previous hypotheses, nor systematically varying from previous hypotheses in the direction of greater appropriateness.

There is quite a difference, of course, between a trial and error of muscle movements à la Guthrie and a trial and error of hypotheses each of which programs a series of choice-point responses. However, this difference would not in particular disturb Ashby, for in his more elaborate model, the multistable system, the activated uniselector under some conditions would switch in and out whole subsystems, and under

other conditions would emit variant settings within a subsystem, as described above.

But Ashby's respectable company on this point leaves many problems yet unsolved.

For one thing, his Homeostat cannot learn. It can solve problems, as described, but it starts from scratch on any new problem, it cannot store old solutions for reuse when the same problem repeats itself. Ashby recognizes this limitation (1, p. 133) although he claims that in showing persistent change in the face of experience, the rudiments of learning and memory are present (1, p. 19). But as W. Grey Walter (25) has pointed out, Homeostat is more plant than animal, and a plant can be claimed to show memory at this level. To quote Ashby:

> If the homeostat adapts to an environment *A* and then to an environment *B*, and is then returned to *A* again, it has no adaptation immediately ready, for its old adaptation was destroyed in the readjustments to *B*; it does not even start with a tendency to adapt more quickly than before; its second adaptation to *A* takes place as though its first adaptation had never occurred (1, p. 133).

Later he suggests that the model can be made to fit better.

> It is convenient for simplicity, to assume here that the first reaction is no longer able to be disrupted by subsequent events. The assumption demands little . . . and there is, in fact, some evidence to suggest that, in the mammalian brain, step-functions that were once labile may become fixed (1, pp. 186–187).

This device, coupled with a multiplicity of subsystems, is in part to describe learning. To the present writer the effort fails because Homeostat or Multistat has no way of recognizing environment *A* or environment *B*, or distinguishing between them. And such recognition is essential to learning. Homeostat and Multistat are blind. They have no sense receptors. The only input is the disturbance that initiates their blind floundering.

Another difficulty. Even were we to grant that Ashby's machine showed consistent purpose in continuing trial and error behavior until a need was reduced, it would still not be showing the full character of adaptive behavior as we know it. For adaptive, purposive behavior demonstrates itself not only in the learning of habits, but also in the execution of habits. It is characteristic of the learned response in many if not all habits that the response is executed in consistency with environmental objects and places, that the executed response has to be described in terms of minor ends or minor purposes (e.g., 4, 14, 23). The typical response in the execution of a habit is a getting to a choice point, a getting off of a grid, a touching of a lever, a picking up of a food bit. Such responses do not run off blindly, but are flexibly guided in terms of the location of environmental objects and places. And if there is error, the next response or portion of response represents a correction of the error, it has a directionality based upon the character of the error. Were this not so, our habits would have little use for us—there would be only one locus for each habit from which it could be effectively applied. But such guided, aimed, goaled responses contradict Ashby's basic gimmick for adaptation—that of the selective survival of random emissions. At this impasse, one might be tempted to place Ashby's book on the shelf. However, a possibility of resolution remains.

Ashby's animal is, as we have said, blind, deaf, anosmic, insensate. It is all motoric. Its only feedback from the environment is frustration. We can imagine that the first primitive protozoa were thus—perhaps Jennings' *Stentor*, which plays such an important role in Ashby's book. Such an organism learns of an obstacle by not being able to go any further. It distinguishes food from non-food by being nourished or poisoned when it ingests things. Its motor efforts provide its only sensory input. We can imagine a course of evolution for such an organism in which special appendages would evolve for—or some legs be used for—explorations of the environment prior to a total body commitment. Such "feelers" would reduce the energy of exploration by total body trial and error, through substituting an abbreviated, truncated, vicarious trial and error. In this fashion, sense organs would evolve as "trial" motor organs—as scouts and pathfinders radiating

out around the main body and eliminating the expensive waste of involving the main body in trial and error. For the guided, adaptive responses characterizing the habits of complicated animals, sense organs are needed. Ashby's model of adaptation can be sustained if we locate in the operation of these sense organs the random response emission, and provide a mechanism by which the major motor efforts of the organism can profit from the substituting sensory trial and error.

Aiming machines are characteristically gifted with sensory processes that can be interpreted as substitute motor exploration. The radar screen searches the sky with a random or unguided pattern. When the beam hits a plane, a gun is then appropriately aimed. The random trial and error of inexpensive electromagnetic wave impulses has been substituted for a trial and error of expensive projectiles. Note that the automata of Walter (25) and others, though less elegantly derived, have blindly scanning photo cells, which, combined with locomotor devices, give them a free and purposive lifelike character lacking in Homeostat.

If we can accept random sensory exploration as a guiding substitute for random motor exploration, Ashby's general mechanism of adaptation may perhaps still hold, and in addition we may have generated a fruitful way of conceptualizing the role of sensory processes in the execution of learned behavior. The implications seem manifold and challenging. The "Stimulus-Response" symmetry in our thinking is abandoned, much as Dewey predicted it would be (6), and "S" and "R" become parallel, intersubstitutable aspects of the same process. Thought can be seen as developing from perception as a still more vicarious substitute trial and error, in which even the supporting presence of environmental objects is no longer needed. And the Pavlovian model for learning appears as most appropriate for a blind organism, neglecting the major capacities of the perceptual mechanisms.

REFERENCES

1. Ashby, W. R. *Design for a brain*. New York: Wiley, 1952.
2. Bertalanffy, L. v. The theory of open systems in physics and biology. *Science*, 1950, 111, 23–28.
3. Brunswik, E. *The conceptual framework of psychology*. Vol. I, Number 10, International Encyclopedia of Unified Science. Chicago: University of Chicago Press, 1952.
4. Campbell, D. T. The operational delineation of "what is learned" via the transposition experiment. *Psychol. Rev.*, 1954, 61, 167–174.
5. Cannon, W. B. *The wisdom of the body*. New York: Norton, 1932.
6. Dewey, J. The reflex arc concept in psychology. *Psychol. Rev.*, 1896, 3, 358–370.
7. Guthrie, E. R. Association and the law of effect. *Psychol. Rev.*, 1940, 47, 127–148.
8. Guthrie, E. R. *The psychology of learning*. (Rev. Ed.) New York: Harper, 1952.
9. Guthrie, E. R., & Horton, G. P. *Cats in a puzzle box*. New York: Rinehart, 1946.
10. Hilgard, E. R. *Theories of learning*. New York: Appleton-Century-Crofts, 1948.
11. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
12. Krechevsky, I. "Hypotheses" in rats. *Psychol. Rev.*, 1932, 39, 516–532.
13. Krechevsky, I. A study of the continuity of the problem solving process. *Psychol. Rev.*, 1938, 45, 107–133.
14. Lashley, K. S. *Brain mechanisms and intelligence*. Chicago: University of Chicago Press, 1929.
15. Lashley, K. S. An examination of the "continuity theory" as applied to discriminative learning. *J. gen. Psychol.*, 1942, 26, 241–265.
16. Lewin, K. *Field theory in social science*. New York: Harper, 1951.
17. Maier, N. R. F. The specific processes constituting the learning function. *Psychol. Rev.*, 1939, 46, 241–252.
18. Nissen, H. W. Description of the learned response in discrimination behavior. *Psychol. Rev.*, 1950, 57, 121–131.
19. Skinner, B. F. *The behavior of organisms*. New York: D. Appleton-Century, 1938.
20. Thorndike, E. L. *The psychology of learning*. New York: Teachers College, 1913.
21. Tolman, E. C. Behaviorism and purpose. *J. Philos.*, 1925, 22, 36–41.
22. Tolman, E. C. A behavioristic theory of ideas. *Psychol. Rev.*, 1926, 33, 352–369.
23. Tolman, E. C. *Purposive behavior in animals and men*. New York: Century, 1932.
24. Tolman, E. C. *Collected papers in psychology*. Berkeley: University of California Press, 1951.
25. Walter, W. G. *The living brain*. New York: Norton, 1953.

A CALCULUS FOR ETHICS: A THEORY OF THE STRUCTURE OF VALUE. PART I

by Nicholas M. Smith, Jr.

Operations Research Office, The Johns Hopkins University

This is the first of two articles dealing with a theory of values based upon a stochastic model of a system. Value is associated with the state of a system. The value of the *state* is defined as the sum of values of future states multiplied respectively by the probability of reaching them. The value of a *thing* is defined as the partial derivative of the state value function, and from this definition a method is derived for computing cost of things and effectiveness of processes which obtain them. Value postulates are developed for immortal and mortal systems, and various means of resolving conflicts in decision among values are considered. The process of *concrescence* is suggested as a synthesis of conflicting values which is logically similar to the scientific method.

I. INTRODUCTION[1]

A. Description of what theory is about

IT is appropriate that a statement be made concerning what this theory is and what it is not about. The theory shall concern itself primarily with the formalistic aspects of values—aspects of values in their most general application. It is contended that the theory should be comprehensive in order to provide a rational setting for a detailed consideration of value experiences, whatever their source may be. In particular, formalistic congruences will be drawn between this theoretical framework and those of the other rational disciplines. A general thesis of this discussion will be that there exists a most general rational form from which the form of all types of rationalization, scientific theories, ethical systems, etc., may be derived by specialization and by the addition of the substance of particular experience. This principle, together with a discussion of its potential value, will be discussed more fully in the latter portions of this paper. It is a theory that is still in the process of development. This material has been presented for the most part in various publications and presentations (2, 5, 6, 18, 19, 20, 21, 22, 23, 27).

[1] The author wishes to acknowledge the very great help and encouragement given by the sympathetic interests of Dr. W. E. Cushen, who has ever been willing to listen and to criticize a continuing revision of these ideas, who suggested the use of Whitehead's term "concresence" for the dialectic process and the term "subjective normalization" for the process presented as a postulate in the section on Relativity, and who also has given considerable aid in correcting the proof of this paper. Original co-investigators were Dr. Franklin C. Brooks, who computed many of the illustrative examples in attrition systems; Dr. Stanley S. Walters, who has made a rigorous investigation of the fundamental value theorem; Prof. David H. Blackwell, who has worked out numerous examples of connecting value theory to game theory and who has investigated the asymptotic behavior of the value equation under certain limiting conditions. I am indebted to Dr. George E. Harmes for the suggestion of the name "analogical conformity" for the process described under that title. The author is also indebted to numerous suggestions that have arisen in the course of oral presentations of this material. This investigation has been conducted as part of the approved work program of the Operations Research Office which operates under contract with the Department of the Army. Miss Jean B. McConville of the ORO Editorial Division has been very helpful in the preparation of this manuscript for publication.

B. Semantics

It is desirable to clarify the semantics employed here in the description of things. Things are considered *in situ*. That is, the *thing* is not to be considered apart from its environment. In a certain sense a particular thing is to be defined as a particular way of classifying the universe. For example, in defining myself it may be stated that the universe can be divided into two parts: my insides and my outsides. Both components are required to describe what is meant by myself as a personality. A different boundary in the same universe may define another thing. An automobile is not a vehicle abstracted and floating in outer space, although such an automobile would consist of all the properties that constitute its "insides." To fully define the automobile one must also describe its "outsides." An automobile consists of (besides a vehicle) a system of road networks, maps describing these roads, a series of service stations, repair and maintenance facilities, and, not at all least, a driver. The word "automobile," for example, in a European setting, does not mean the same thing as the same word in an American setting. The outsides are different, and this in turn reflects on the designs of the insides.

In the same sense, a description of value of a commodity refers to that commodity in its environment. Value is not a property of the commodity in abstract, but of the commodity *in situ*.

C. Illustrative example

That the value of a dollar bill is not an absolute quantity but depends more on the circumstance in which it is exchanged can be illustrated very simply by noting the comparative effort a starving man would make to obtain a dollar bill compared to that which a rich man would make to get the same dollar bill. To the starving man the dollar bill represents something to eat, and therefore represents survival. It could in fact represent his whole chance to live, whereas to the rich man the dollar bill represents an increment toward the achievement of some further goal not necessarily of tantamount importance to his survival.

Attention becomes directed not to the particular things or items in the system such as the man, the dollar bill, or the automobile, but to the system as a whole. It shall be seen that values are fundamentally to be associated with the state or condition of the entire system, and not necessarily with some linear combinations of the things comprising it. The discussion that follows introduces a mathematical model based upon the theory of stochastic processes. In a paper of this length it will not be possible to go very deeply into the mathematical theorems and proofs. These proofs will be outlined, but their detailed exposition must necessarily await a comprehensive volume on this subject.

II. STOCHASTIC MODEL

A. Definitions

It is necessary to define a technical language. For this technical language we shall draw upon the terminology of physics and mathematics as it pertains to stochastic processes (1, 4, 7, 8, 28). A stochastic process is to be distinguished from a deterministic process. If event A is always followed by event B, which is always followed by C, etc. (see Figure 1a), each change occurring with a probability of unity, then once the

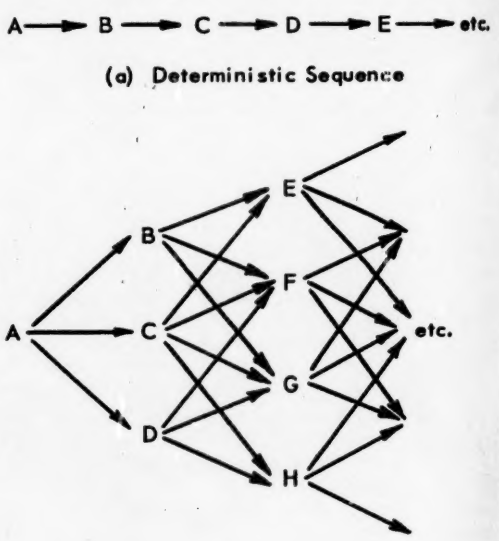


FIG. 1. Illustration of Deterministic and Stochastic Series of Events.

condition A is reached the complete chain of events is fully determined. On the other hand (see Figure 1b), if event A may be followed by either B or C, then the event is said to be "stochastic." A series of such events, each followed by many alternatives, is described as a stochastic process. If the branching ratios are known, the system is *stochastic definite*. The real world can be more nearly described in terms of a stochastic model than in terms of the deterministic model.

1. State. The condition of a system at any moment in which it is not undergoing change is described as a "state" of the system. A state may be defined as a collection of numbers:

$$x = (x_1, x_2, x_3, x_4, \cdots x_i) \qquad [1]$$

where the state x of a system is described by the set of numbers, x_1, x_2, etc., and a number of components are included *sufficient adequately to describe the system with respect to the particular problems faced*. For example, in a system comprised of two opposing military forces, the state of the system might consist of a listing of the numbers of separate men and weapons on both sides. Such a collection of numbers, together with the transition probabilities, comprises a model. This collection will be represented here by any of the lower case letters in the latter part of the alphabet.

2. Transition and transition probability. Any change in the system is considered to occur discretely. For example, in the array of numbers x_1 to x_i, x_3 might possibly represent the numbers of friendly tanks in the opposing military forces. One of these tanks might be destroyed by the enemy. The system is said to have undergone a transition:

$$(x_1, x_2, x_3, x_4, \cdots x_i) \rightarrow (x_1, x_2, x_3 - 1, x_4, \cdots x_i). \qquad [2]$$

That is, the change that occurs may be in only one of the array of numbers (i.e., the "coordinates") that describe the state. The word "event" may be used synonymously to express such a transition. The probability that any particular transition occurs is defined as a "transition probability" and is represented by the symbol $P(x, s; y, t)$, $t > s$. Special consideration must be given as $t \rightarrow s$ or as $s \rightarrow t$. (See, for example, Arley, Ref. 1, p. 25.) This notation is to be read as a probability that the system initially at state x at time t will be found in state y at time s.

3. System. A "system" may be defined as a set of all states that are connected by nonzero transition probabilities. This definition, although adequate for the first part of this discussion, will have to be modified later. In the first part of this discussion only attrition systems will be explored. An attrition system is one in which the vector components always decrease—as would occur, for example, in a battle that is isolated from the sources of supply of the participants.

4. Special states. There are special kinds of states, or sequence of states, to which reference may occasionally be made. Mention already has been made of the "deterministic sequence" where one state follows another with probability of unity—that is, there is no branching. There may be cyclic states in the system; i.e., a deterministic sequence that folds back on itself to produce a continuously recycling path through the same set of states. In simple attrition systems primary interest occurs in the "trapped" states. A trapped state may be defined as a state for which there exists no chance to escape; i.e., there are finite transition probabilities to the state from others, but one for which the transition probabilities away from the state are zero. A system once arriving in the condition of a trapped state undergoes no more change. It becomes "trapped" in that particular condition. For example, such a system could be a gambling game. The state of this system could be expressed by the amount of money held by one player. The system will become trapped if either player wins all the money. When a system becomes trapped—there being no further change—the game ends.

B. State probability

Consider a system initially in a state w at time s, and at some later time u in a state y. It may have arrived at the state y along many different paths. Consider a path that

traverses state x at an intermediate time t. The probability that the system went from w to y along the particular path that led to x is the product of the probability that it first moves from w to x and then from x to y in the proper temporal sequence. If now these conditional probabilities are summed for all possible paths there results the well-known Chapman-Kolmogorov equation (12, 13):

$$P(w,s;y,u) = \sum_x P(w,s;x,t)P(x,t;y,u), \qquad u > t > s. \tag{3}$$

where

$$\sum_y (P(w,s;y,u) = 1,$$

and $$0 \leq P(w,s;y,u) \leq 1.$$

1. Forward state equation. There are two distinct ways in which the Chapman-Kolmogorov equation may be regarded. First, it may be considered that the system is known to have been initially in state w at time s. Equation 3 is then employed to determine the probability that the system may be found at state y at some later time u. That is, the equation so interpreted is given a known condition in the present to predict the probability of the future.

In this usage the equation is known as the "forward" state equation, or more simply as the forward equation. The subscript zero $(_0)$ will be employed to designate a fixed known state, the forward equation being written as

$$P(w_0, s_0; y, u) = \sum_x P(w_0, s_0; x, t)P(x,t;y,u), \qquad u > t > s_0. \tag{4}$$

2. Backward state equation. On the other hand the final condition of a system may be assumed to be known. That is, it may be known that at time u the system was in state y. The Chapman-Kolmogorov Equation 3 then may be employed to determine that the probability that the system initially in the field point (x, t) traverses the fixed point y_0 at time u_0. This is the so-called "backward" or "adjoint" equation and is designated here with the zero subscripts as follows:

$$P(w, s; y_0, u_0) = \sum_x P(w, s; x, t)P(x, t; y_0, u_0), \qquad u_0 > t > s. \tag{5}$$

The adjoint equation can always be interpreted as a value function. This interpretation arises as Equation 5 represents the contribution that state w makes toward the achievement of the final state y_0. That is, the backward probability gives the *importance* of state w to this final condition.[2]

3. The value equation. Although Equation 5 is fundamentally an *importance* equation it is not in the form in which the value equation appears. Let the value associated with the state w at time s be denoted by $Q(w, s)$. The fractional contribution to the value function $q(w, s, y_0)$ by a particular trapped state y_0 is to be obtained in an ultimate sense. This contribution is defined as proportional to the limit of the adjoint probability function (Equation 5) as u_0, the time to reach the final state, moves out to infinity. Or, in other words, this contribution is defined as the probability for any time that the system will pass from state w at time s ultimately to state y_0. The contribution to the state value is defined to be proportional to this limit:

$$q(w, s, y_0) = \operatorname*{Lim}_{u_0 \to \infty} P(w, s; y_0, u_0). \tag{6}$$

For the time being, attention will be limited to those systems that are certain ultimately to come to rest in a trapped state. This may be regarded as an assumption, or conversely as a constraint placed upon the present considerations. Simple attrition systems, if left to themselves, would have such an ultimate fate. If in the gambling game previously mentioned the players gamble long enough, one or the other will surely win all of the money. In a battle fought to the last man where other factors do not intervene to stop

[2] This is related to the equation of retrodiction. See Ref. 28, Part III. To get retrodiction the transition probabilities must be a probability measure over the antecedent states, not the consequent states as expressed in Equation 3.

the battle, the fight would continue until there were no survivors on one or the other of the participating sides. Since, by hypothesis, only attrition systems are considered for the moment, there need be no concern about the system's avoidance of its ultimate fate.

Having assumed that y_0 is a trapped state, that is that

$$\sum_{y_0} \operatorname{Lim}_{u_0 \to \infty} P(w, s; y_0, u_0) = 1, \qquad [7]$$

enough time has elapsed so that the system is in one or the other of the possible trapped states. Then, an *arbitrary* weighting of the value of the trapped state, $\mu(y_0)$, leads to the value definition:

$$\begin{aligned} Q(w, s) &= \sum_{y_0} q(w, s, y_0) \\ &= \sum_{y_0} \mu(y_0) \{ \operatorname{Lim}_{u_0 \to \infty} P(w, s; y_0, u_0) \}. \end{aligned} \qquad [8]$$

Equation 8 says that the value of the state w at time s is the sum over all trapped states of the product of an arbitrary weighting of each trapped state multiplied by the probability that the system ultimately reaches it. For example, if there are two kinds of trapped states, those desirable y_d and those undesirable y_u, and if furthermore the desirable trapped states are given a weight of unity and the undesirable trapped states a weight of zero, then the value $Q(w, s)$ becomes identically equal to the probability that the system ultimately comes to rest in any one of the desirable states:

$$\begin{aligned} &\text{if} \quad \mu(y_d) = 1, \\ &\qquad \mu(y_u) = 0, \\ &\text{then} \\ &Q(w, s) = \sum_{y_0 = y_d} \{ \operatorname{Lim}_{u \to \infty} P(w, s; y_0, u_0) \}. \end{aligned} \qquad [9]$$

The value function, as it appears in Equation 8, is more often expressed as the relationship between the value of one state at the reference time to the values of the states that may be reached at some later time. For example, referring to Equation 5, let us ask what the value is of the state x at time t. In conformity with the definition of Equation 8 this becomes:

$$Q(x, t) = \sum_{y_0} \mu(y_0) \operatorname{Lim}_{u \to \infty} P(x, t; y_0, u_0). \qquad [10]$$

Referring back to the adjoint Equation 5, if the time u_0 is allowed to go to infinity and the definitions of Equations 8 and 10 are applied, one obtains:

$$\begin{aligned} Q(w, s) &= \sum_{x} P(w, s; x, t) Q(x, t), \\ & t > s. \end{aligned} \qquad [11]$$

Equation 11 states that the value of the present state w is the average (or expected) value of a future state x, or in other words, is equal to the product of the probability of transition from state w at time s to state x at time t, with the value of the future state x at time t. Equation 11 is the value equation.

C. The meaning of value

From the preceding discussions of the definition of value and its derivation from the adjoint probability, its fundamental meaning becomes clear. In Equation 9 it is seen that the value of a present state w is linked to the probability of ultimate success, that is, of ultimately reaching desired goals. In general, since the arbitrary weighting functions $\mu(y_0)$ have been introduced, the value to be associated with a particular state at a particular time becomes an arbitrary linear combination of the probability of ultimate arrival of the trapped states. It is a usual convention to consider that the more desirable states are valued more positively. Since this weighting is arbitrary, at least in this state of the description of value one may choose the weighting assumed in Equation 9, or, if he prefers symmetry, may assign the undesirable value states a value of -1. It is said that the value placed upon the trapped states is arbitrary. It is arbitrary in the sense that any set of values $\mu(y_0)$ permit a solution of Equation 11. Equation 11 also calls attention to the continued future reference or estimate of values. The values of the states today are to be determined by the values of the states reached tomorrow. Those in turn are the values of the states that may be reached the day after tomorrow, and so on apparently ad infinitum. In order to obtain a solution to Equation 11 it is necessary to define some boundary conditions under

which the values are assumed to be known. In the present instance, this boundary was assumed to exist on a set of trapped states.

In contradistinction, the forward probability function (Equation 4) has a unique solution if it is known that the state is initially at state w_0 at time s_0. One is essentially looking backward to determine the probability for the future in the forward equation. Conversely, in the value equation, one looks to the future to determine the value of the present and the past.

1. Values in nonattrition systems. It is not necessary that the definition of value be restricted solely to attrition systems, nor to systems to which the ultimate attainment of a trapped condition is certain. Value may be defined in terms of specific endpoints that may occur at time u_0. That is, for Equation 8 one replaces Equation 12:

$$Q(w, s; y_0, u_0) = \sum_{y_0} \mu(y_0, u_0) P(w, s; y_0, u_0), \quad [12]$$

$$u_0 > s.$$

In Equation 12 the summation is taken over all the possible states that could be reached at time u_0. Equation 11 again follows from this more general definition. In either case an arbitrary constant or arbitrary function has been introduced to the value equation. Throughout this discussion this arbitrariness will be emphasized. Value determinations continually refer to future states, and those in turn to states further into the future, and so on, until either an endpoint is reached or until the state is so far into the future that prediction becomes impossible. At either of these points rational operation ceases and one must resort to an intuitive designation. These are essentially postulates of the value structure. Later in the discussion it will be suggested that there are more fundamental and more absolute principles that constrain the intuitive approach to the assignment of ultimate values; the application of these principles, however, does not lead to a unique set of values—that is, to a unique solution to the value equations. There is a natural repugnance felt by some readers that ultimate values are to be considered arbitrary and intuitive and they prefer to consider them otherwise. This possibility requires further development of the theory in its present rigid framework and will be discussed in a section Part II on the epistemological problem in value theory. After showing the relationship between decision criteria and value theory, a more detailed discussion of value postulates will be made.

D. Values and decision procedures

At this point, value theory may be linked with decision theory. Decisions rest not only on the probabilities of outcomes, but upon the values of these outcomes, or in terms of the present state w at the present time s there is some requirement to maximize the expected value—all other factors being considered. Consider the simplest situation first.

1. The elementary criterion of decision. Consider now that an individual has some control of a situation existing in a stochastic system. That is, he may take certain actions i that affect the transition probability. The transition probabilities thus depend upon the action taken and are denoted

$$P^{(i)}(w, s; x, t),$$

where the superscript i indicates a particular action taken. The value to be associated with the state w at the time s thus becomes Equation 13:

$$Q^{(i)}(w, s) = \sum_{x} P^{(i)}(w, s; x, t) Q(x, t), \quad [13]$$

$$t > s.$$

In terms, then, of a set of values $Q(x, t)$ over the future states x, there now exists not a single value associated with state w at the present time s, but a set of values—one associated with each of the possible courses of action i. One may now introduce an elementary criterion of decision (sometimes referred to as a game against nature), namely, one selects that course of action i for which its associated value is a maximum:

$$Q(w, s) = \operatorname{Max}_{i} \{Q^{(i)}(w, s)\}. \quad [14]$$

This decision criterion simultaneously produces a unique course of action and at the same time selects a single value from among a set of values associated with each course of action. Equation 14 may be written as follows:

$$Q(w, s) = \underset{i}{D} \{Q^{(i)}(w, s)\}, \qquad [15]$$

where the particular instructions to select the Q having the maximum values are replaced by the more general instruction symbolized by D_i ; namely, to apply a particular decision criterion. The relationship between a single value $Q(w, s)$ on the left of Equation 15 and decision criteria that may be applied to the right of Equation 15 constitutes the subject of the theory of decision processes, and is not the subject of this discussion. Equation 15 may be generalized to cover the two-or-more-person game situations.

2. Statistical criterion in n-person games. It frequently occurs that not one, but two or more players control the magnitude of the transition probabilities. If two players have such independent control, the situation is called a two-person game and Equation 15 becomes Equation 16:

$$\left.\begin{aligned} Q(w, s) &= \underset{i,j}{D} \{Q^{(i,j)}(w, s)\}, \\ Q^{(i,j)}(w, s) &= \sum_x P^{(i,j)}(w, s; x, t) Q(x, t) \\ & \qquad t > s. \end{aligned}\right\} \qquad [16]$$

The second part of Equation 16 indicates that not a single set of values is associated with the transition probabilities in the future state x, but a matrix of values—one value for each combination of actions (i, j) taken by the two players. The decision criterion D determines the course of action i for one player, j for the second player, and at the same time a unique value is associated with the matrix of values. A decision criterion (3, 24, 26) commonly applied to the two-person game is the so-called minimax principle. Another might be the so-called principle of minimum regret. In the remainder of the discussion of value theory it will be assumed that an appropriate decision criterion is imposed to determine courses of action that permit the replacement of the matrix of values with a unique value, and considerations of application of Equation 16 will be taken for granted and not explicitly mentioned. The generalization of Equation 16 for high-ordered games is obvious.

3. A fundamental invariancy. The above definitions, equations, and conditions result in a fundamental invariancy that is of interest in applications to game theory. Consider a system that passed through a fixed initial point, that is state w_0 at time s_0 , and then at a later time u_0 passes through a fixed endpoint y_0 . Both the forward and the backward probability functions are defined for the time interval between s_0 and u_0 ; that is, for a state x that is reached in a time t in this interval. The state x at time t may be considered a variable field point (see Figure 2).

The forward function described the probability that the system will be in state x at time t having started from the fixed initial point. The backward probability expresses the importance of the variable field point to the fixed endpoint. The product of the forward and the backward probabilities is the weighted importance of the variable field point. The sum of this product over all possible intermediate variable field points constitutes what is called the inner scalar product of the forward and backward probability functions—or as we shall refer to it, simply as the scalar product. For conditions where the Chapman-Kolmogorov Equation 3 holds, it is very simple to demonstrate that the scalar product defined in the interval

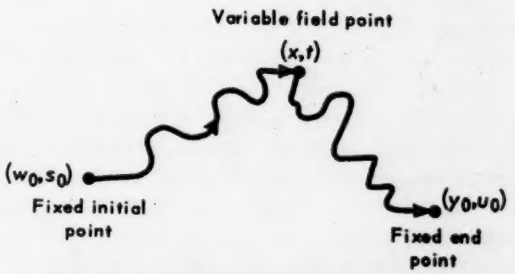


FIG. 2. Stochastic Process with Fixed Initial and End Points.

$u_0 - s_0$ is independent of the time. That is, in this interval it is an invariant:

$$\left.\begin{aligned} &\sum_x P(w_0, s_0; x, t)P(x, t; y_0, u_0) \\ &= \sum_{x'} P(w_0, s_0, x', t')P(x', t'; y_0, u_0), \\ &= P(w_0, s_0; y_0, u_0), \end{aligned}\right\} \quad [17]$$

where

$$u_0 > t > s_0,$$
$$u_0 > t' > s_0.$$

The proof is a simple one and is omitted. A simple extension of this theorem leads to the invariancy of the scalar product of the state forward probability with the state value function:

$$\left.\begin{aligned} &\sum_x P(w_0, s_0; x, t)Q(x, t; y_0, u_0) \\ &= \sum_{x'} P(w_0, s_0; x', t')Q(x', t'; y_0, u_0), \\ &\qquad u_0 > t > s_0, \\ &\qquad u_0 > t' > s_0, \end{aligned}\right\} \quad [18]$$

or, for attrition systems

$$\left.\begin{aligned} &\sum_x P(w_0, s_0; x, t)Q(x, t) \\ &\qquad = \sum_{x'} P(w_0, s_0; x', t')Q(x', t'), \\ &\qquad = Q(w_0, s_0), t > s. \end{aligned}\right\}$$

The significance of this theorem is that values so defined, coupled with a particular decision criterion described above, permit even attrition games to be treated as zero-sum games. A nonzero-sum game is to be defined in terms of the inapplicability of the Chapman-Kolmogorov equation. Such a situation would arise if values are associated with the particular paths that the system takes in going from the initial to the final point rather than upon these final points alone. For example, in a gambling game where the house takes a percentage of each pot, the value of reaching the winning state certainly depends upon the path taken. In such a system the Chapman-Kolmogorov equation does not apply; it is a nonzero-sum system.

III. FUNDAMENTAL VALUE POSTULATES

The definition of the value function as expressed in Equation 8 or as generalized in Equation 12 has incorporated into it an arbitrary function. The insertion of this arbitrary function is essential since the adjoint probability function alone is only the probability with which a given state will lead to a certain outcome. In the adjoint function there is no means of differentiating between desirable and undesirable outcomes. There is furthermore no means of differentiating between degrees of desirability, as for example, in Equation 9, where desirable goals are assigned a value of plus unity and undesirable goals a value of zero. As has been stated, such assignment is completely arbitrary and may be considered a postulate of the value system. The arbitrary function introduced into the definition of value permits an intuitive weighting of the final states of the system. If many of us can agree on our basic value postulates and if furthermore the decisions are made consistently with respect to values determined from these value postulates then there can be a source of common or of uniform action. Or the viewpoint of the cultural anthropologist may be taken and the behavioral patterns of a culture observed in terms of decisions made. Next a search is made to find a simple set of value postulates to which these actions appear to be consistent. Obviously, varying degrees of success will be encountered in testing these postulates with which a given culture appears to adhere; they must be advanced tentatively, and tested against direct observation. It is in this spirit then that the subject of the assignment of fundamental value postulates is here developed.

A. Fundamental theorem

It is illuminating to explore values in attrition systems as imposed by some extreme value postulates, and it is desirable to add one further restriction on the characteristics of the attrition system. For the time being, at least, the discussion will be confined to time-wise linear systems. A stochastic system is said to be time-linear if the transition probabilities are not functions of the absolute time, but only of the time difference.

That is:

$$P(w, s; x, t) \rightarrow P(w, x; t - s), \quad t > s. \qquad [19]$$

If this condition is applied to the limit expressed in Equation 7 it is seen at once that the resultant function is independent of the time. Thus the value associated in such a linear attrition system depends only on the state of the system and not on the time coordinate, and the value Equation 11 may be written simply as follows:

$$\left.\begin{aligned} Q(w) &= \sum_x P(w, x; t - s)Q(x), \\ & t > s. \end{aligned}\right\} \qquad [20]$$

As already indicated in Equation 18 the summation on the right-hand side of Equation 20 is independent of the duration of time $t - s$ during which the system might evolve from state w to state x. A fundamental theorem may now be stated: *There is a unique value function $Q(w, s)$, satisfying Equation 11 above, which agrees with an arbitrary assignment of values $\mu(y_0)$ assigned to the trapped states, and that furthermore this value is independent of time when the transition probability $P(w, s; x, t)$ depends only on $t - s$.* The uniqueness proof is not included herein although it evolves from a straightforward procedure (23). The fundamental theorem may be generalized to nonattrition systems of the type for which the value definition of Equation 12 applies, provided again that the transition probability is time independent and furthermore provided that the postulate for value of any boundary state is a constant in time.

1. An example of the solution of the value equation in an attrition system. Solutions of the value equation in simple attrition systems have been studied for a variety of conditions. One of these will be described here. Consider that task forces of two opposing nations—RED and BLUE—are roaming around in a large area. Occasionally two opposing forces will meet and engage in a battle of annihilation. For simplicity assume that the elements of the task forces are identical. In a battle between a particular RED and BLUE force, the number of BLUE elements will be designated by the letter m, and the RED forces by n. The pair of numbers, (m, n), $0 \leq m \leq m_0$, $0 \leq n \leq n_0$, describe the state of the system comprising the battle between these two specific forces. The history of the battle may be represented on a phase plot in two dimensions with the m axis the abscissa and the n axis the ordinate (see Figure 3). Each permitted discrete combination of nonnegative numbers m and n represent a possible state of the system. The state (m_0, n_0) is the initial state. It shall be assumed furthermore that the nature of the engagement is such that Lanchester's Square Law (14) applies.

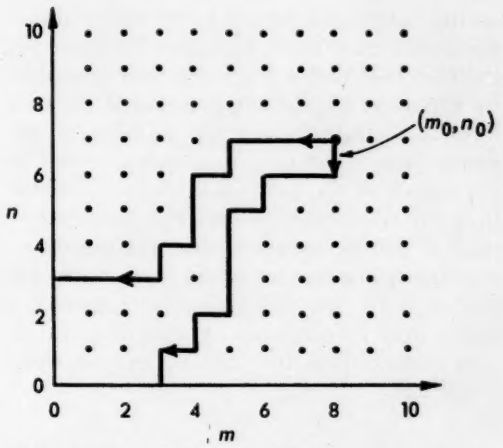


FIG. 3. Two-Dimensional Representation of Battles of Attrition.

The Lanchester relations refer to a deterministic description of such a battle. The rate of loss of BLUE forces is proportional to the size of the RED forces, and vice versa:

$$\left.\begin{aligned} \dot{m} &= -k_{mn}n, \\ \dot{n} &= -k_{nm}m. \end{aligned}\right\} \qquad [21]$$

These simple differential equations are referred to as the square law because if one imposes a condition of parity, namely, that the percentage loss rates on either side are equal,

$$\frac{\dot{m}}{m} = \frac{\dot{n}}{n}, \qquad [22]$$

or in mathematical language the logarithmic derivatives are equal, there results:

$$k_{nm}m_0^2 = k_{mn}n_0^2 \qquad [23]$$

as the conditions for expected parity under the initial conditions. Equation 23 expresses the fact that under the conditions described by Equation 21, the effectiveness of the task force as determined by the number of elements that constitute it is proportional to the square of the number of elements rather than to their first power. For simplicity's sake it will be assumed that the constants of proportionality are equal. The stochastic analogue to the deterministic Lanchester battle may be obtained by assigning transition probabilities for the loss of one BLUE element:

$$q = \frac{n}{m + n}, \qquad [24]$$

that is, for a transition from the state (m, n) to the state $(m - 1, n)$. Similarly, the transition from the state (m, n) to the state $(m, n - 1)$ is given by:

$$p = \frac{m}{m + n}. \qquad [25]$$

When one side or the other becomes totally destroyed in this battle, the battle ends. Thus the axes of the plot in Figure 3 represent the trapped states of the system. The states $(m, 0)$ are those states for which there are BLUE survivors and no RED survivors, and are therefore winning trapped states for BLUE, and conversely the states $(0, n)$ on the ordinate axis represent the winning trapped states for RED, or the losing trapped states for BLUE. The values will be

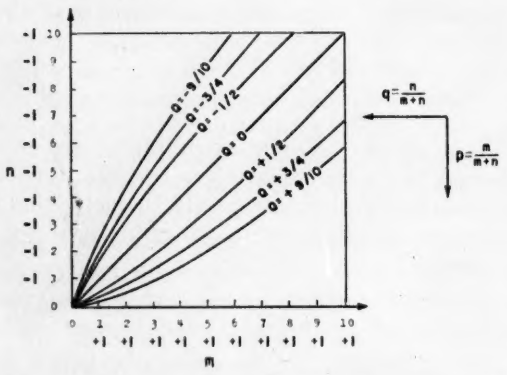


FIG. 4. Ultimate Values in a Stochastic Analogue of the Lanchester Square Law. (Source, Reference 22)

assigned from BLUE's viewpoint. Two cases will be considered.

a. *Case 1.* Assume that this battle is for the control of a strategic position. If BLUE can hold, reinforcing task forces will have time to arrive to save the entire situation. Should BLUE lose this battle, RED will overrun the position—gaining an advantage and probably winning the war. Thus the prize won by winning the battle is many orders of magnitude greater in value than the value of the surviving elements. The value of the elements may be ignored in relationship to the value of the prize. Under these conditions, it should be possible to come to a satisfactory agreement concerning the value postulates to be assigned to the trapped states of this battle. By convention the more desirable states are considered to have values more positive. Since the prize of this battle is the entire national integrity or the life of the system from BLUE's viewpoint, it seems reasonable under the circumstances to assign each one of his winning states an equal value and for simplicity's sake the value $+1$ is chosen. By symmetry each of the losing trapped states is assigned the value -1. It is of course of no consequence whether these values are chosen to be zero or -1. In the present instance, -1 affords a symmetrical presentation. Had the value zero been chosen for the losing trapped states then as already seen from Equation 9 the value of the intermediate states would be exactly equal to the probability for BLUE's winning the battle.

The value Equation 11 is solved essentially in reverse. The state (1,1) is first selected—the value of this state being given by Equation 11 in terms of the ultimate postulates on the trapped state. This solution will permit a solution of the value for the state (2,1), etc. by iteration until the complete solution within the framework (m_0, n_0) is obtained. On the basis of these values at these discrete points, contours of equal value may be erected and Figure 4 represents equal value contours in this situation. The line of parity is the line for which the state value is equal to zero. This might also be called the condition of indifference.

A system at parity has equal probability

that either side will win. The deterministic description of the battle would be that it proceeded down the line of parity until both sides were completely annihilated. In the stochastic system this particular path has the least probability. As a curious fact the paths having greatest probability are those for which one side is completely decimated and the other side is completely unharmed. That is, the line of parity is the line of instability. A particular path in this phase plot may be called a history of a system, and it may be observed that the expected histories from points near parity have the greatest uncertainty as to the final outcome. As the initial conditions are farther and farther away from the line of parity, the outcome becomes more nearly deterministic. Thus the deterministic Lanchester Differential Equations 21 may give perfectly adequate description of the system far from the line of parity. Close to the line of parity, however, a stochastic model is required.

b. *Case 2.* Now consider another extreme viewpoint for the assignment of values on the trapped states of the systems. Suppose in a war of annihilation between the RED and BLUE task forces the battle between the particular RED and BLUE task forces illustrated here has only marginal significance. One is concerned then with the value of the surviving element. If the war is to continue, the surviving elements are to comprise a smaller task force that will engage other enemy forces whenever it meets them. Hence, the value of the surviving task force, it shall be assumed, is to be weighted in proportion to the probability of its winning its next battle, also assuming that winning a war still remains of paramount value with respect to any particular intrinsic value of the individual elements. It will be shown in Section III C that the methodology of marginal analysis requires the state value to be separable into two functions—one a function of friendly forces alone and the other a function of enemy forces alone. That is, a solution to the value equation is sought such that

$$Q(m, n) = S(m) - S(n). \qquad [26]$$

Moreover, it should be expected that the separate functions $S(m)$, $S(n)$ are themselves

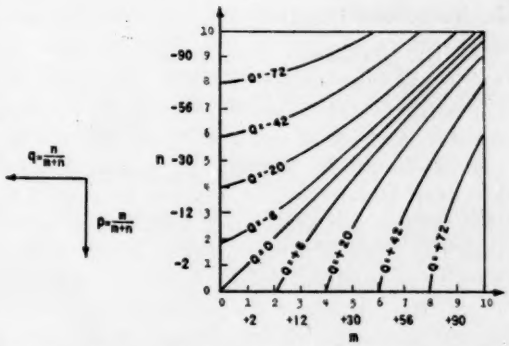


FIG. 5. Marginal Values in a Stochastic Analogue of Lanchester's Square Law. (Source, Reference 22)

proportional to the square of the number of surviving elements in conformity with the characteristics of the Lanchester Square Law.

If the value $m(m + 1)$ is assigned to the winning trapped states $(m, 0)$ and $-n(n + 1)$ is assigned to the losing trapped states $(0, n)$, then the value of an arbitrary state (m, n) is

$$Q(m, n) = m(m + 1) - n(n + 1). \qquad [27]$$

The solutions giving contours of constant value for this case appear in Figure 5. It is to be noted that the value function given in Equation 27 is the only nontrivial one that has the form $S(m) - S(n)$. Thus, it appears that in the stochastic analogue the conditions for parity that correspond to the Lanchester square relationship are:

$$m_0(m_0 + 1) = n_0(n_0 + 1). \qquad [28]$$

A third case can be visualized. Assume that this battle between two task forces would be the only battle in which these particular elements could be brought to bear; that they would have no opportunity to enter into a second battle. In this case it might be argued that the intrinsic value of the survivors would be the controlling postulate, and that the trapped state values would be assigned proportional to the surviving elements. It appears, then, that there are many situations, particularly in battles of attrition, in which reasonable and agreeable assignment of trapped state values can be made.

B. Value of a thing

In the discussion thus far the concept of value has been associated with the state of the entire system. This is properly the fundamental and most general way in which values should be viewed. It is customary, however, in our past experience, to consider that values are associated with things and that the value of the system as a whole is a linear combination of the values of things. It was the purpose of the opening discussions on the semantics of things and the illustrative example of the value of a dollar to emphasize that this viewpoint is in general terms erroneous. Only in very special cases can the value of the system be considered to be composed of a linear sum of the values of the things that compose the system. The fundamental notion of value does not depend upon the things themselves, but upon states of things.

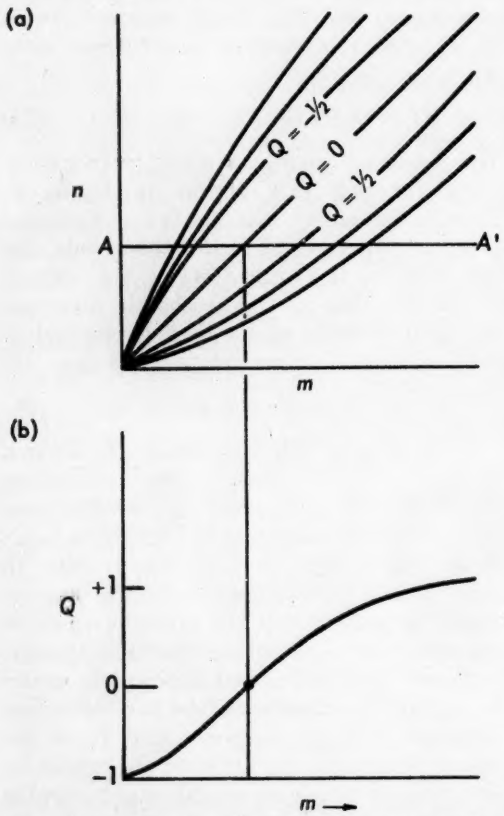


FIG. 6. Illustration of a Cut Through the Value Function.

The value of a thing is here defined as the difference in state values of a system whose states differ by the thing under consideration. This definition is equivalent to defining the value of a thing in terms of the partial derivative of the state value function with respect to the coordinate appropriate to the particular thing under consideration. In discrete terminology the value of the thing is given by the ratio $\Delta Q/\Delta x_i$ where Q, the state value function, is a function of the state x at time t and where x is the vector quantity composed of the commodities $(x_1, x_2, x_3, \cdots x_i)$. The value of a thing is thus defined in the neighborhood of the state of the system, and is in general not a constant quantity but a function of the state of the system. The task force whose values are illustrated in Figure 3 may be used to demonstrate this point.

Consider Figure 6a where the plane AA' passes perpendicularly through the value function for a constant value of n. Consider the intersection of this plane with the value function as plotted in Figure 6b. The value function Q for $m = 0$ has a value of -1, it then rises monotonically to the value $+1$ with increasing m. It has a point of inflection at the point $Q = 0$ corresponding to the condition of parity. At the point $Q = 0$ the probability that BLUE will win this battle is equal to ½ and the outcome is completely uncertain. This function, of course, is defined only for discrete values of m and n and is illustrated here as a continuous function only for matters of convenience. By definition, the derivative of the function illustrated in Figure 6b is the value of a single element of the BLUE task force as defined with respect to the immediate neighborhood.

This derivative is illustrated in Figure 7. The value of a single element of m for $m = 0$ is equal to zero, it rises to a maximum value that occurs at the condition for parity—that is, the condition at which the outcome of the battle is uncertain—and then falls to zero. The portion of the curve in Figure 7 labeled as 1 can be taken to illustrate the Law of Diminishing Returns. Whenever the probability of success is greater than ½ and therefore success is reasonably certain, the

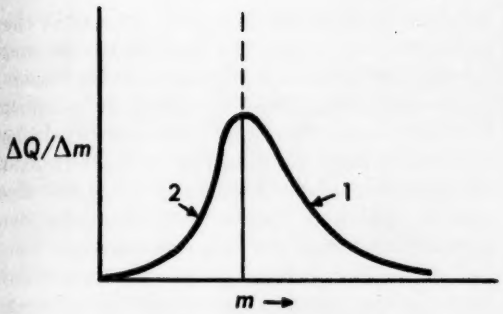


FIG. 7. The Value of a Thing.
(The Derivative of the Value Function)

addition of a single element to the task force adds less value than was added by the preceding quantity, and so on. That is, the value of the things diminish as they are added to the system. Note again that the condition for the applicability of a law of diminishing returns is that the probability of success is greater than ½; i.e., that success is expected. Thus one expects the law of diminishing returns to apply to a system whose chance of survival is somewhat greater than even.

On the other hand, in respect to that portion of the curve in Figure 7 labeled 2 one might propose a Law of Increasing Returns. Here an element added to the task force has a value greater than that added by the preceding element. Values are increasing. Note also that this condition exists whenever the values are below parity; i.e., for conditions under which the survival of the system has less than an even chance. It then applies to very desperate situations. It is reasonable to suppose that one seldom hears this law expressed because it applies to systems whose death is imminent and that do not exist long enough for consideration. There is an old saw to the effect that "a drowning man will grasp at a straw"—essentially a statement of the Law of Increasing Returns, applying to conditions where death is highly probable and situations are very desperate.

The Law of Increasing Returns is the basis of exploitation of desperate people by those who are not in such dire circumstances. This effect may be considered to be true in general. As the system approaches the undesirable state the values of things takes on an augmented and unusual level and the decisions made in the system may appear as highly erratic and irrational to an external observer who himself has relative security and plenty.

C. Marginal analysis

Most analytical procedures presently in use for application to decision purposes employed by operations analysts and other professional groups in the terminology of this theory are marginal analyses. A marginal analysis is defined as that pertaining to a projected change in a system of such a small magnitude that the change in state value produced as the result of a given action can be considered infinitesimal compared to the difference in state value between the least desirable and the most desirable states. In other words, the action contemplated affects the ultimate goals with respect to the system in an incremental or marginal manner. A second condition imposed is that the transition probabilities be homogeneous in time. Under these conditions the state value function $Q(w)$, possessing the proper analytical properties may be expanded in a Taylor series in the phase space defined by the coordinates of the state w around that state w as an origin. Such a Taylor expansion written in difference notation would have the following form:[3]

$$\Delta Q(w) = \sum_i \frac{\Delta Q(w)}{\Delta w_i} \Delta w_i + \sum_i \sum_j \frac{\Delta Q(w)}{\Delta w_i w_j} \Delta w_i \Delta w_j + \text{etc. higher-ordered terms.} \qquad [29]$$

In the expansion of Equation 29, the terms $\frac{\Delta Q(w)}{\Delta w_i}$ represent the value of the i^{th} thing evaluated in the neighborhood of the state w, and the terms Δw_i represent the number of things of the i^{th} type consumed by the projected action. The higher-ordered derivatives contain cross-product terms that represent the change in the value of the i^{th} com-

[3] The second-order terms may be used to define a "mutual" value distinct from a "self" value (first-order terms) and may be used to correct first-order marginal analysis.

modity as affected by the change in the value of the j^{th} commodity, etc. By the conditions assumed for the marginal situation all terms of higher order than the first linear set may be neglected, yielding:

$$\Delta Q(w) = \sum_i \frac{\Delta Q(w)}{\Delta w_i} \Delta w_i \qquad [30]$$

as the fundamental value equation applied to marginal situations. Consider now the application of Equation 30 to a battle or game between two contestants. The summation that represents the sum of the product of the value of each type of thing by the number of things consumed may be divided into two classes: the friendly things consumed and the enemy things destroyed. Or, in an economic situation, it could represent raw products consumed and a second class representing final items produced. This leads to:

$$\Delta Q(w) = \underbrace{\sum_i \frac{\Delta Q(w)}{\Delta w_i} \Delta w_i}_{\substack{\text{friendly} \\ \text{weapons} \\ \text{consumed}}} + \underbrace{\sum_i \frac{\Delta Q(w)}{\Delta w_i} \Delta w_i}_{\substack{\text{enemy} \\ \text{weapons} \\ \text{destroyed}}} . \qquad [31]$$

Notice that if higher-ordered terms had been included in this expansion, it would not have been possible to separate the change in state value into two terms associated with these two classes of commodities: friendly weapons consumed, enemy weapons destroyed. *The negative of the first term is defined as the cost associated with the projected action and the second term is defined as the effectiveness of the projected action. Only in a marginal situation does the concept of cost and effectiveness as separable quantities have any meaning.*

As an example of the breakdown of the concept of cost and effectiveness, consider an analysis of decisions pertaining to the establishment of an outpost. It shall be assumed that the mission of the outpost is to delay an enemy's advance for a period of, say, 10 days, or a sufficient time to enable the mobilization of forces in the interior. The decision problems concern questions of magnitude of resources and personnel to be placed in the outpost. The quantities of such personnel and material in a sense might be considered the cost of the outpost. But assume that an outpost of insufficient strength is established, that the enemy overruns the outpost in less than the specified time and catches the interior unprepared and conquers it. In such a situation the cost of the outpost is to be reckoned not only in the commodities that went into its establishment, but also the losses incurred depending upon the degree to which it fulfills its mission. In the hypothetical case the loss was extreme. The entire national integrity was lost, hence the cost of the outpost in which insufficient resources were placed becomes the loss of the entire national values, and not just the resources placed into it. That is, the cost cannot be considered independently of the effectiveness of the outpost. Since cost and effectiveness are not separable quantities in such a situation, they cease to be useful concepts. Such a breakdown occurs when nonmarginal values are considered. In the illustration cited, the values instead of being marginal were total in that they affected the ultimate goals of the system.

Assume that the decision problem has to do with a selection between various ways of accomplishing the marginal task. In comparing the various actions, one may either keep the cost constant and compare the effectiveness, or keep the effectiveness constant and vary the cost. The former is the usual methodology employed by operations analysts in the field where the concern is with the maximization of results with a given weapons system. The weapons system —that is the cost—remains fixed and is compared to the effect produced. On the other hand, in problems involving an initial selection of a weapons system, a given military requirement is selected as a common measure of effectiveness and the cost of various weapons systems required to produce this constant effectiveness is compared, in order to decide between them.

1. An example: intrinsic values and the allocation of common costs in a marginal analysis (22). Value theory has been successfully employed to give solutions to two very puzzling problems in cost analysis. Suppose that the same organization or facility is employed in two different types of missions to produce two kinds of products, say type A and B. There are costs in the organization

and/or facility that accrue irrespective of the quantity of A or B produced. These are so-called common charges. What fraction of the common charges is to be assigned to the cost of production A, and what fraction to the cost of production B? With no further restrictions than those stated, this allocation is perfectly arbitrary.

For example, in the TVA systems the cost of the dams and lakes is chargeable to three missions: flood control, rehabilitation of land, and the production of electricity. How much of the common cost of the dam construction is to be charged to the production of electricity? It may be argued that the production of electricity is purely a by-product and that none of the fixed common charges should be charged to the production of electricity. The net result would be a minimum cost of electricity delivered to the consumer. On the other hand, a competitor to the TVA system might argue that the commercial electricity produced should be required to completely pay for the TVA system. Such an allocation of fixed charges would result in a maximum cost to the consumer of electricity produced. In the actual case, neither extreme is taken.

Another very important situation is encountered at all levels of the decision process and may be included in the illustrative example above, and this concerns the desirability or necessity of the introduction of *intrinsic value* as an *imponderable* into the decision process. In order to set up the problem, let it be assumed that there exists a strategic material involved in the production of A that is not utilized in the production of B. This strategic material is not only relatively costly, but it is in short supply; there is not enough material to produce all the commodities of type A which could be utilized. But the effectiveness of A far exceeds that of B although they are used for the same purpose. The cost-effectiveness ratio for type A is in fact much less than that of type B. Thus, in order to supply the requirement, it is necessary to produce all the commodities of type A as permitted by the quantity of strategic material and to satisfy the remainder of the commodities by producing that of type B. The decision problem that is faced has to do with the methods of production of commodity type A. This is essentially a suboptimization (9). There are various ways that this can be done. What is the cheapest method?

One might, for example, select a method by which the production of a single item of type A, irrespective of the larger problem, was cheaper than by other methods. Yet its effectiveness was so reduced that it materially increases the requirement for the commodity of the less efficient type B. Hence, the suboptimization of the method of producing type A cannot be considered independently of the cost of producing the commodity of type B. *The introduction of the intrinsic value of the strategic material type A into the suboptimization of the production of the commodity A is a device by which the suboptimization of production of commodity A may be conducted in such a fashion that total optimization results simultaneously.* In general it may be stated that whenever intrinsic values or imponderables must be introduced into a decision process it is indicative that a suboptimization is being conducted. But the real problem requires an enlargement in scope of factors considered. One may introduce intrinsic values purposely in order to conduct a suboptimization with a maximum degree of awareness for the total optimization problem.

The following notations will be employed:

n_{ai} = number of commodities of type A produced in the i^{th} manner;

a_i = cost of production of one unit of type A by the i^{th} method, including the cost of producing the strategic raw material consumed;

n_{bi} = number of commodities of type B required to fulfill over-all requirements of A, when A is produced by the i^{th} method (B is always produced by same conventional method);

b = cost of production of each commodity of type B;

$f_{ai}^{\dagger}$ = fixed costs directly traceable to i^{th} method of producing commodity of type A;

$f_b^{\dagger}$ = fixed costs directly traceable to production of commodity of type B;

f_i = common costs incurred when A is produced by i^{th} method;

f_{ai} = fraction of common costs charged to production of A by i^{th} method;

f_{bi} = fraction of common costs charged to pro-

duction of B when A is produced by i^{th} method;

v_{ai} = intrinsic value of a unit of strategic material consumed in the production of one item of type A by the i^{th} method;

E = total effectiveness, a constant requirement;

E_{Ai} = effectiveness accomplished by all commodities of type A, when A is produced by the i^{th} method;

$E_{\ i}$ = effectiveness accomplished by all commodities of type B, when A is produced by the i^{th} method;

S_i = relative adequacy of strategic stockpile with respect to meeting requirements entirely by commodity of type A when produced by the i^{th} method, $S_i = E_{ai}/E$;

U = total stockpile of strategic materials;

u_i = amount of strategic material used in a single item of type A when A is produced by the i^{th} method.

This problem is illustrated in Figure 8. The closed contours represent the respective costs. At the top are the traceable noncommon variable costs that are proportional to the number of items produced, in the second layer are the traceable noncommon fixed costs, and in the third layer the common fixed costs that must be arbitrarily divided and allocated among method A and method B. To these have been added fictitious intrinsic value costs, which total zero.

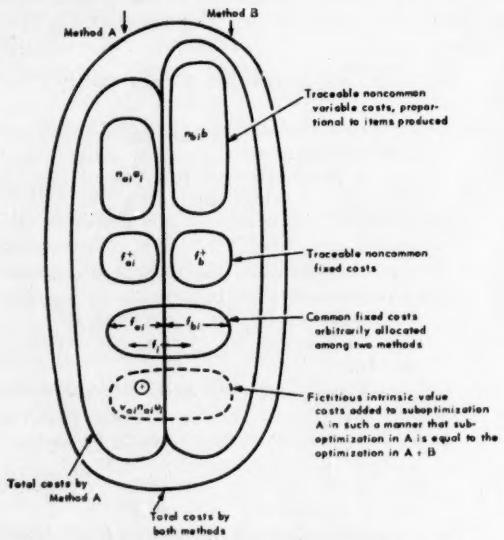

Method A
Method B
Traceable noncommon variable costs, proportional to items produced
Traceable noncommon fixed costs
Common fixed costs arbitrarily allocated among two methods
Fictitious intrinsic value costs added to suboptimization A in such a manner that suboptimization in A is equal to the optimization in A + B
Total costs by Method A
Total costs by both methods

FIG. 8. Illustration of Intrinsic Value Cost, Common Cost, and Traceable Variable Costs in Suboptimal and Total Optimal Analyses. (Source, Reference 21)

The concern here is the portion of these added to the cost of production of commodity A. There exists a unique solution to the assignment of intrinsic value of the scarce strategic material consumed in the production of A, and to the division of the common fixed costs only if one imposes certain reasonable side constraints. Otherwise, the solution is completely arbitrary.

Because of limitations of space, the theorem giving the solution to this problem will be stated without proof:

A unique assignment of intrinsic value of a strategic material and the allocation of common costs may be made if:

(*a*) it is required that the suboptimization of production of commodity A,—that is, the selection of the i^{th} method—be made in such a fashion that the same selection would be made if the combined production of A and B (that is, the total optimization) were considered;

(*b*) the intrinsic value costs over and above the actual cost that is assigned to the strategic material must go to zero whenever the strategic material is no longer scarce (that is, whenever the total requirements can be supplied by the commodity A); and

(*c*) the allocation of common costs is made according to the same rule whether or not an intrinsic value is introduced and conversely, the rule for assigning intrinsic values is the same whether or not the allocation of common fixed cost has to be faced.

Under the condition of these constraints the common fixed cost is to be divided in proportion to the ratio of total effectiveness accomplished by commodities A to that accomplished by commodities B:

$$\frac{f_{ai}}{f_{bi}} = \frac{E_{Ai}}{E_{Bi}} \qquad [32]$$

$$f_{ai} + f_{bi} = f_i$$

and the intrinsic value assigned to the strategic material is proportional to the cost of replacing commodity A by commodity B multiplied by the fractional inadequacy of the stockpile of strategic material:

$$v_{ai} = \frac{1}{u_i} S_i(1 - S_i)\left[\frac{E_A}{E_b} b_i - a_i\right] \qquad [33a]$$

$$= \left(1 - \frac{E_A}{E}\right)\left(\frac{D_{bi} - D_{ai}}{U}\right) \qquad [33b]$$

The theorem stated in Equation 32 yields a powerful tool for the study of cultural values. In the example of the TVA dams, the employment of this theorem states that if the common costs are to be divided among

its three missions in such a fashion that total optimization for the nation results, then this division is to be made in proportion to the value of the three separate missions to the nation. What is the value of flood control with respect to rehabilitation of land? With respect to the production of electric power to the nation? Such national values are largely unknown. When facing such a decision for the first time the theorem has no usefulness a priori. It does, however, have a usefulness a posteriori. One may use the actual decision reached as to the allocation of common charges and essentially utilize the theorem in reverse; namely, that past decisions imply certain values. Thus, the actual manner of allocating common charges by inference determines the national values that in effect are held. Similar problems arise in the setting of rates for interstate transportation. What fraction of the cost of erection and operation of railroads should be allocated to the cost of passenger transportation, and what fraction should be charged to the cost of freight transportation? The theorem would state that this allocation of charges should be in proportion to the national value of freight to passenger transportation. In an actual case, these values may be unknown. One uses the theorem in reverse and the actual allocations that are made are used to determine the national values that exist in effect.

2. Cost-effectiveness ratios. One sees first of all that the concept of intrinsic value that would appear essentially as an imponderable in the suboptimization of systems A_i was introduced in such a fashion that the problem of larger scope was in effect considered. If the initial attack to the problem had been from the viewpoint of the problem of larger scope, there would have been no necessity for the introduction of the concept of intrinsic value of the scarce commodity. Thus it appears that whenever the analyst feels that intrinsic values and imponderables must be introduced into his problem to give reasonable weighting to the solution, it is indicative that the study was not conducted on a sufficiently broad scope. The use of intrinsic values approximately determined by problems of larger scope is a technique that can be employed in suboptimal problems under conditions for which there is no free market place to which appeal could be made by observation to determine the level of the intrinsic value.

One also sees in this methodology a technique for continually broadening the scope of an analysis with respect to more and more diversified uses for a strategic material as well as for the introduction of other factors.

It is probably good technique to carry out each suboptimization with due consideration of the next broader degree of optimization. In this manner one can proceed by iteration between studies of adjacent areas to successfully broader states of the analysis of very complex and broad problems. For example, assume that the stockpile of strategic material considered in Equation 33 is only the portion of the total stockpile allocated to partially meet the required objective given in this particular example; that there are other uses for the strategic material in which it is in general in short supply. Each problem may be treated separately and the intrinsic value of the strategic material determined in each specific category of use. The stockpiles allocated among these varied uses can then be adjusted until the intrinsic value of the strategic material is the same for each possible use. *When such a condition is reached the allocation is optimal.*

Indeed, it may be stated as a general result that the value-to-cost ratio of the expenditure of resources, or as more commonly expressed, the effectiveness-to-cost ratio of expenditure of resources, is in the optimal equilibrium system equal for all commodities involved. This means, then, that in situations in which the total effectiveness is held constant, the effectiveness-cost ratios may not be used as an index leading to preferred actions. Instead, the difference between the effectiveness and cost must be considered. This law may be proven independently or may be considered as the corollary to the theorem on intrinsic values. It extends not only into economic matters, but also into tactical weapons systems. In a stable tactical weapons system the effectiveness-to-cost ratio of any one component of the system is equal to that of any other. Thus, if an

entirely new weapon that has an unusually high effectiveness-to-cost ratio is introduced into the system, it may be predicted that the environment will react to the introduction of this weapon until equilibrium is again established.

Another use for the intrinsic value concept is that it permits a more rapid convergence in the iteration of solutions of problems in adjacent areas such that each problem is optimized with a reasonable (but not completely detailed) consideration for the solution of the adjacent problem.

D. Values in immortal and mortal systems

In general, real systems are not as simple as the ones described above. In the first place, real systems are not always sharply defined. The original definition of system stated that it was the set of all states that were connected by nonzero transition probabilities. This definition is strictly applicable only to highly simplified models that have some of the characteristics abstracted from the world of experience. Real systems are never completely closed. They are generally embedded in an environment to which they are coupled in various degrees, as Miller (16) and his associates observe in referring to them as "open systems." One must continually remind himself that his rational operations pertain to a *model* having characteristics abstracted from the wide range of his percepts. The decision problem itself must generally act as a guide for the formation of a model. In principle it is desired to abstract all of those qualities that affect the decision importantly and if at all possible, omit all of those qualities that do not affect the decision importantly (20).

It is also found that our knowledge of the system is often incomplete. Since value judgments constantly refer to future states, it is not always possible to be fully cognizant of all the possible future states of the system. Who in 1935, for example, projecting his viewpoint to 1945, would have predicted the existence of atomic weapons? Developments constantly introduce new conditions that cannot be anticipated into real systems. New developments not only in technology, but also in methodologies, in language, and in theory, constantly enlarge the volume of the known universe in a set of possible states. Some systems must be considered in which the winning states do not have significance. This is particularly true in looking at national values as the nation develops from one point to another. During a war the system considered is a set of two nations at war. When a war is ended that system ends, but the nation—at least the winning nation—continues. At the beginning of the war it is not unreasonable to make decisions according to the principle that as long as the war is won any final state is to be considered to be of the same value as any other. However, as the end of the war approaches and it is seen that it will probably be won, more and more consideration is given to the state of the nation after the war. It is then customary to assign intrinsic values to the surviving element after a battle of attrition. Yet, for these intrinsic values one must look further into the future to the utilization of these elements for individual goals in future situations. What is the value to be assigned to the winning state, for example, in which a nation is left so weak that it could readily be overcome by a third one? Hence it appears that the values of the final states of one system have to do with the success or failure of a system formed of the surviving elements of that particular system, and so on. In the case of a nation, national integrity itself may be considered to constitute a prime value of a system. During a war the nation is a subsystem of the combined nations at war. During periods of peace it may form coalitions with other nations, and thus appear as a subsystem in many other different supersystems. Hence in every system there appears to be a residuum of values that refers to the next future problem of overcoming stresses, etc. ad infinitum.

It shall be attempted in this section to state some postulates that are believed to be basic and fundamental to all systems. It should constantly be kept in mind that these postulates are intuitive in character; that it is not the attempt of the author to say that systems should adopt these postulates as their own. It is the attempt of the author on observing systems, to postulate that they

acted *as if* they had these postulates in mind. Whether or not these postulates satisfactorily explain the actions of individuals and systems must be left to experimental observation, historical study, and to the reader's own intuitive approach to the same problems. The statements of value postulates to follow may not necessarily be the best ones, and most certainly not the only ones; it is believed that they contain some of the most elemental ones. They are presented here in order to stimulate the investigation into this important problem and to serve as a basis for discussion.

1. Survival. The peacetime values of a nation are not so sharply defined as those it possesses when engaged in a war. This may be true, however, only because the possible stresses are numerous and the particular stress for which it should prepare itself is uncertain. Over the long run, it appears that a basic and fundamental value is associated with survival. Postulate 1 is therefore proposed:

Postulate 1: *All systems strive to perpetuate themselves.*

In terms of this postulate alone and within certain qualifications to be discussed under the topic of "Mortal Systems," the value to be assigned to the surviving states is a linear function of the survival probability with respect to the stresses imposed upon the system. For simplicity one may adopt the convention that the death state of the system is assigned a zero value. Thus in terms of Postulate 1 the value to be assigned any other state in a system is the survival probability.

Systems are often stressed simultaneously from many causes—both external and internal. For example, it may be possible for a government to make decisions concerning national action that would enhance its probability of winning a potential war with an enemy. The consequences of these decisions might being about internal stresses within a nation; that is, within the values adopted by the subsystems within it, such that these subsystems would repudiate the authority of the government and cause it to meet death from internal stresses. Thus, in assigning values to future states of real systems, one must consider the time-wise application of all the possible stresses.

It is not the purpose here to go too deeply into time-dependent value theory. We shall assume, in the notation below, that conditions exist (See III A) under which values can be expressed independently of the time variable and shall conclude as the result of Postulate 1 that the value associated with state w can be expressed:

$$Q(w) = \sum_x P(w, x, t - s) \prod_i Q_i(x), \quad [34]$$

where $Q_i(x)$ represents the survival probability with respect to the i^{th} stress. The equation reads, in words, that the value of the state w is given by the product of the survival probabilities of the state x with respect to all possible future stresses, multiplied by the transition probability from the state w to state x in time $t - s$.

Equation 34 successfully transfers the problem of value determination from the present state w to future states x. It would be necessary to reapply Equation 34 at the future states x, referring to even more remotely future states y, etc, the series never being finished. In the case of the simple attrition model the very happy condition existed that enabled the assumption that the system would ultimately come to rest either in a winning or a losing or a draw situation; i.e., it would ultimately and surely become trapped. It is necessary to add another postulate before successful solution of Equation 34 can be achieved. This postulate must take the place of the assumption of ultimate trapping in the case of the attrition system.

2. Unpredictability. Not only do real systems differ from idealized model systems in that one may not be aware of all the possible future states of the system, but also as one projects himself further and further into the future the ability to estimate transition probabilities diminishes. Transition probabilities may be used to express the condition of ignorance as well as a bona fide branching probability in the presence of complete knowledge. Thus a rational action in the absence of any information might be to assume that all outcomes have an equal probability. Should a commander attack the enemy on its right flank or its left flank?

In the absence of any information whatsoever concerning the enemy's tactical doctrine or his position or deployment, it would be reasonable to suppose or to act as if the probability of winning the battle was the same whichever flank were attacked. This leads then to a second postulate that will permit the termination of the apparently endless process of referring to future and more remotely future situations.

Postulate 2.: *There exists a limiting state probability distribution for any system as it is considered further and further into the future. This limiting distribution is brought about by the increasing unpredictability of the system states and the increased variance of the transition probabilities. A corollary to this postulate, when combined with Postulate 1, states that there exists a limiting state value as the conditions for unpredictability are approached.*

This postulate is stated rather vaguely, primarily because the consequences of increasing unpredictability have not been thoroughly investigated and must remain a problem for future research. It has been observed by von Neumann (25) and Watanabe (28) that the antecedent problem of retrodiction has a diffuseness, increasing as more and more remotely prior times are considered because of the degradation of information. Postulate 2 here states that there is an equally increasing diffuseness introduced into the consequent problem (prediction), which increases with increase of time ahead of the present. The best decision model can be expected to diverge increasingly from a real situation as time progresses. This introduces a diffuseness in prediction over and above that introduced by the stochastic nature of the process. Rationality can exist only in a small island in time and space around the present—fading into uncertainty and intuition in the past and future. The implication of Postulate 2 as stated is that the completely unpredictable system may be considered in effect a random walk of the phase point in the space defined by the system which is uniform except for the existence of a set of death states. Some very simple systems have been considered in which it appeared that the probability distribution was uniform over all possible states and that the probability of survival was linear with respect to the state density. Assuming this conclusion to be correct, the limiting value under conditions of unpredictability would be a value of indifference.

The effect of increasing the variance of the estimate of transition probabilities, even for a system whose states are known, is to move the expected transition probability from any particular value to that corresponding to uniform probability distributed among the possible immediate outcomes. One is concerned here in physical terminology with the entropy of information concerning the future system. The effect of the increase-of-information entropy will be to consider the system to approach more and more that of the completely random system. More investigation is required to clarify this concept. The net effect of the general idea introduced by Postulate 2 is that the continued mathematical induction toward future and more future states can be terminated. This termination is needed in order to fix the solution of the value equation.

5. Values in mortal systems. Although it was not explicitly stated in the preceding section, it becomes clear that the first two postulates taken alone refer to systems that have a finite chance of survival ad infinitum; that is, they refer to *immortal systems*. What happens to the values in a system that sooner or later must meet a stress that it cannot successfully overcome? An examination of Equation 34 reveals that the value in an immortal system is proportional to the product of the survival probabilities with respect to all the stresses to which it is subjected. *If any one of these survival probabilities is zero the entire product is zero, and the estimated values collapse everywhere to a value of zero.* Indeed, the entire value structure becomes trivial and useless. By implication if values are destroyed, rationality is destroyed, and ability to reach decisions is lost.

If any of the survival probabilities with respect to future stresses is zero, then the system is no longer immortal, but transient, and will be referred to here as a *mortal system*. The simple attrition systems previously considered were such systems. It is proposed

therefore, to consider some of the consequences of the assumption of mortality. Under what conditions can values exist? Postulate 3 is therefore proposed:

Postulate 3: *Ultimate death of a system is inevitable.*

Individuals, nations, even whole cultures go through a cycle of adolescence, maturity, decline, and ultimate destruction. The sun and the stars go through their own cycle of birth, evolution, maturity, decline, and death. Entire galaxies of stars have their own history of evolution. Atoms, even fundamental particles, are subject to change. It appears that all things that are identifiable as systems are transient in character. It is the purpose here to discuss the consequences of the postulate of immortality. Therefore, such religious doctrine as concerns the existence of an immortal state and the indestructibility of the human soul will not be discussed, although these exist as very important postulates that lead to the establishment of values. In particular, they lead to value systems of the type previously discussed. What is the consequence of Postulate 3? Can an ethical system exist in the presence of the inevitability of ultimate destruction?

Throughout this discussion, the systems considered have been limited to those for which the Chapman-Kolmogorov Equation holds, and in particular, the concept of value has been associated with states of the system. It is a consequence of the postulate of mortality that the definition of value must be broadened or no value structure may exist. Postulate 4 is therefore proposed:

Postulate 4: *The value of a state of a mortal system depends in general upon the path taken to the ultimate, inevitable death.*

In general then, one is led in mortal systems to propose non-homogeneous value systems. A few simple cases will be considered and it will be seen that in portions of the history of mortal systems and for certain simple conditions a return to homogeneous values is possible.

a. *Case 1. Specific desired goals.* A simple manner of weighting the paths (the histories) of the system from a particular state to the death states exists whenever there is a set of

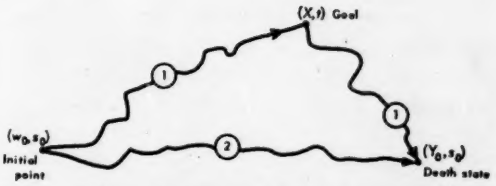


FIG. 9. Schematic Representation of a Stochastic System Containing a Desired Goal and a Death State.

Path 1 is a desired path and is weighted as value of +1.

Path 2 is an undesired path and is weighted as zero.

goals that are desired to be reached before the death of a system. A path that does not lead through a desired goal might be given, for example, a value of zero, whereas a path that leads through any of the desired goals might be given a positive value—say unity. Figure 9 illustrates the desirable and undesirable paths for the case of a single goal. The probability that the system proceeds from the state w_0 at time s_0 to the desired goal X at time t is given by the probability kernel $P(w_0, s_0; X, t)$. If the state X is a *null* state (i.e., if the mean time of the system once at X to return to X is infinite—the goal is reached only once in the lifetime of the system) and if, furthermore, there is no concern over the amount of time consumed in reaching a goal and it is valued equally independently of the time required to reach it, the value associated with the state w at time s is given by Equation 35:

$$Q(w, s) = \sum_X \int_{t=s}^{\infty} P(w, s; X, t)\, dt \quad [35]$$

where the summation is included to extend considerations for paths through other goals equally weighted. Or, if the other goals are not equally weighted, an arbitrary weighting function may be introduced as in[4]

$$Q(w, s) = \sum_X \mu(X) \int_{t=s}^{\infty} P(w, s; X, t)\, dt. \quad [36]$$

The weighting of the desired goal, of course, may be made a function of time. A simple

[4] If X is an ergodic state, i.e., if the reoccurrence time is finite, Equations 35 and 36 must be replaced by more complicated expressions. (See Feller, Ref. 8, p. 320.)

weighting would be that imposed by a deadline after which the attainment of the desired goal is of no value. In this case the integral in Equation 36 would be cut off at the deadline time. Note the similarity between Equation 36 and the former definition of value given in Equation 8. The limiting process in Equation 8 has been replaced by the integral in Equation 36, the fundamental difference being that in Equation 8 the system was assumed to terminate when the trapped state was reached, whereas in Equation 36 the system simply passes through the desired states X. If the definition of Equation 36 is now combined with the Chapman-Kolmorogov Equation 3, there is obtained again Equation 11, the fundamental value equation. Thus the conclusion is reached that in systems having desired goals that portion of their history that leads to the desired goals—or to the death state—may be considered exactly as it was under the discussion of attrition systems. In particular, if the transition probability is time-linear, then as before, the state value equation results in values that are independent of the time. A desired goal is to be treated exactly as if it were a desired trapped state. Once the system has passed through the desired goal, however, this value system collapses and must be replaced by some other value system appropriate to transient conditions.

It is possible, of course, that as soon as one set of goals is reached a new set of goals may be established and a system may proceed from one set to another, each time establishing a new set of homogeneous values as each set of goals is accomplished, and so on, until the ultimate fate is reached. It must be realized that the value structure changes each time a goal is reached. It is a common experience of individuals and organizations to establish in advance a series of objectives. As the successful attainment of one objective is assured, values gradually shift to those appropriate for the next one.

b. *Case 2. Longevity.* Another simple case of mortal systems to be considered are those whose values are concerned with the maximum possible duration of its existence, in which case the value associated with the state w may be defined to be proportional to the expected lifetime, $l(w, s)$, of the system from that state at that time:

$$(w, s) \sim l(w, s) . \qquad [37]$$

The resulting value structure is not homogeneous. The simple operations that are permitted by value structures previously discussed do not generally apply. One may restore some semblance of homogeneity by selecting goals or a series of goals such that to pass through them would yield a maximum lifetime. The values placed upon the attainment of these goals might be made proportional to this expected life.

c. *Case 3. Value proportional to magnitude of a system coordinate.* Another possible way of postulating values and mortal systems is to let the value of a state be proportional to the magnitude of one of the coordinates associated with that state. One is interested not only in survival and longevity, but in the most "abundant" life. This abundance might be measured in terms of riches, or other possessions amassed. A nation might set values proportional to a standard of living or a gross national product. Such values may or may not lead to a homogeneous system.

d. *Case 4. Preferred decision codes.* One can visualize systems in which values are placed upon those paths that are determined by more abstract constructs—e.g., the saving of "face," "playing the game," by following the socially established rules of conduct, or in general selecting those paths for high evaluation that conform to established moral constructs.

4. Unpredictability in mortal systems. Consider now what happens to values in a mortal system with such complication and diffuseness of predictability that its death, though certain, lies in the unpredictable future. It has been discussed that unpredictability implies in many respects that the system may be considered ultimately as a random walk, perhaps with some absorption states. It is well known that in such systems, if they consist of more than two dimensions, and if furthermore the death states of the system can be contained in a closed surface, that there is a finite probability that the

state of the system may wander away from the neighborhood of the death states and never return. This is equivalent to saying that under conditions of unpredictability such systems, though mortal, may be treated as if they had a finite chance of immortality. This is proposed here by way of suggesting the fifth postulate:

Postulate 5: *If death of the system, though inevitable, is in the completely unpredictable future, values are assigned as if immortality has a finite probability.*

The application of Postulate 5 to mortal systems whose death is not in the predictable future has the effect of restoring the value systems that are determined wholly by survival—that is, value systems appropriate to a doctrine of immortality.

Thus the implication of immortality is found to be inextricably entwined in the value structures of mortal systems. Since the homogeneity of the value system is so important to decision procedures, this may lead naturally to speculations concerning the decision efficiencies as affecting the technological advancement of cultures whose religious doctrines contain the concept of immortality. Fatalistic societies will find the mechanism of decision much more difficult and thus should not be expected to make the "progress" that the nonfatalistic cultures exhibit.

5. Role of the awareness of mortality in quasi-immortal systems. It has been postulated and reasons have been proposed in support that mortal systems for a period of their histories may act as if they were in reality immortal systems; i.e., are quasi-immortal systems. There will come a time, sooner or later, that the inevitable death emerges from the region of unpredictability. If the values of the individual system have been made solely on the presumption of immortality, the values of such a system will be expected to collapse as soon as such emergence occurs and the individual is left without means of making decisions. The same effect occurs for mortal systems that have just passed through a state designated as a desired goal. The state values must be entirely reoriented; and in the case of the predictable approach of death, the reorientation must be from values appropriate to an immortal system to those appropriate to a mortal one. If the concept of immortality is strongly adhered to under these circumstances it may be expected that increasingly desperate and "irrational" (i.e., irrational to one not facing the situation) measures will be taken as death approaches.

In the terms of value theory this effect has a close correlation with the observation of moralists (10, 11) that the well-balanced individual continually has in mind a sense of the tragic (i.e., mortal and transitory) component of life. The application of the consequences of Postulate 5 in the moral realm would emphasize the need for anticipating inevitable shifts in value systems—as the system previously assumed to be immortal approaches its end.

IV. SOURCES OF DISAGREEMENT IN VALUE POSTULATES

In the description of values thus far, it has been stated that the values postulated for the ultimate states of a system are to be considered arbitrary. In some highly simplified examples, values for the ultimate states have been assigned in a manner in which it can be expected general agreement would result. In this section the important sources of disagreement in the assignment of values of ultimate states are to be considered.

A. Differences of intuition

If ultimate values are postulated in a purely intuitive fashion it is reasonable to suppose that the intuition exercised by different individuals will result in different value postulates. Referring to Figure 10a the ellipse represents all the possible states of a system composed of two nations at war. If the system is presently in the state A and an action is contemplated that might move the system to state B, an attempt must be made to evaluate B. In real situations, this evaluation is normally done by very approximate means as to the probability of achieving state B, and the value of state B is estimated on an intuitive basis. Since these estimates depend upon the judgment and experience of the individual and since experience differs this will result in an important

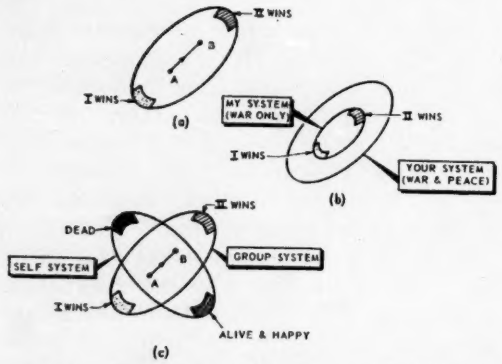


FIG. 10. Schematic Illustration of the Sources of Confusion and Disagreement in Setting Values to the Final States of an Episode. In (*a*) disagreement may result in an estimation of the change of value in moving from state A to state B because the solution is made intuitively. In (*b*) disagreement between the assignment of values of the trapped states may result in differences in areas of closure of systems considered. If I am concerned only with winning the war, no matter what else may happen, then I assign the values of all winning trapped states equal to, say, +1, all losing trapped states −1. On the other hand, you may be considering the reconstruction following the war and wish to value the end of the war proportional to the number of men remaining. In (*c*) is a schematic illustration of a state of conflict in the mind of a tank driver who must choose either to maximize the expectation of his group as a whole, or his personal expectation. (From Reference 22).

source of disagreement. From the theoretical viewpoint however this class of disagreement may be considered a trivial one.

B. Difference of area of closure

Figure 10b illustrates another source of disagreement brought about by the scope of factors included in the consideration of the action. The small ellipse represents the state involved while the state of war exists between two nations. The larger ellipse represents those states in the peaceful reconstruction in addition to the states during the war. One individual may approach the contemplated action as if his sole concern is to win the war, and he values any winning state equal to any other. A second individual concerned not only with winning the war, but also with the peaceful reconstruction, will disagree, and will argue that intrinsic values should be assigned to the survivors.

1. Intrinsic and extrinsic values. This viewpoint leads to an interpretation of the difference between extrinsic and intrinsic values of things. The value of a thing as defined can be a combination of both types of values. If the values of final ultimate states of the system are not weighted proportional to the individual elements surviving, and if all states yielding the prize are weighted equally, then the value of the element in the interior states may be said to be wholly extrinsic; i.e., its value is wholly determined by its utility with respect to the goals of the system. On the other hand, if in addition to the value placed upon the goal, values are added in proportion to the number of surviving elements this value may be considered to be an intrinsic value—intrinsic values with respect to consideration of wider scope. As the scope under consideration becomes greater and greater and approaches the universe of all known factors there may or may not be a residuum of value associated with the individual element (depending upon the postulates of the value system). If it is postulated in this universal state that residual values are still assigned to the individual elements, one has what is normally referred to by philosophers as intrinsic value.

C. Differences associated with states existing simultaneously in several systems

A very important source of differences and uncertainty in the assignment of values occurs whenever the states involved by the given action exist simultaneously in more than one system. Consider a hypothetical example: the dilemma faced by the tank commander in Figure 11.[5] He is facing an

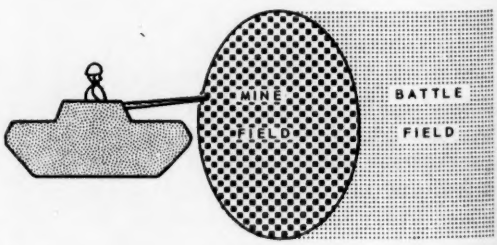


FIG. 11. The Tank Commander's Dilemma.

[5] Taken from a game suggested by D. H. Blackwell.

area in which mines are scattered in unknown places. On the other side of the mine field is a battlefield. The tank commander has been ordered to proceed at all haste through the mine field in time enough to engage the enemy in battle within one hour's time. The mine field is so extensive that by the usual methods of mine detection it would require at least two hours to traverse the mine field safely. This original problem arose in connection with the enemy strategy concerning the layout of the mines and the search strategy pursued by the tank commander.

The question faced here is that of postulating the relative values involved. Three outcomes can be foreseen by the tank commander. Since he has one hour in which to search for mines, he can search for one hour and then proceed without caution through the remainder of the mine field and, arriving there safely, enter into the battle. On the other hand, he might search his way through the entire mine field, consuming so much time that he arrives at the battlefield too late to give aid and assistance to his comrades-in-arms. Lastly, the tank may strike a mine and get blown up. What are the values to be associated with these outcomes? The greatest military value will be assigned to the first outcome—that in which the tank penetrates the mine field safely in time to engage the enemy. If the tank searches too long and does not arrive at the battlefield in time to engage the enemy, the outcome is considered of near zero value since this is an important battle. On the other hand, if the tank gets blown up by the mine field, it is a net loss and cannot be utilized in any future battles.

While thinking this over the tank commander begins to consider the situation in terms of his self-system. Suppose he does get through the mine field safely even though taking his chances in order to arrive at the battlefield on time. He must engage in a very hazardous battle in which there is an appreciable probability that he will become a casualty. So he values this outcome zero in his self-system. In the second place, if he can malinger and take so much time in searching through the mine field, he will not only protect himself against mines, but he will arrive at the battlefield too late to engage in the hazardous battle. This is valued in his self-system as some positive value a'. He will agree (within the frame of his self-system) with the military experts that if his tank gets blown up a net loss results and will value the final outcome as $-b'$. On the basis of the elementary criterion of decision and the military system of values, the tank commander will favor course of action a. He will search in the mine field for as much time as he can spare, take his chances on penetrating the remainder, and hope to get through to assist in the battle. If he is maximizing his self-values, he will malinger, search for mines, and avoid the battle. Thus one course of action in one system of values and a second course of action in the second system of values are indicated. This is shown in Table 1. Where states are common to two or more value systems, sooner or later such a situation is reached. *This is a typical illustration of conflict in decision.* In terms of the ordinary decision criteria there is no way out of this dilemma. Decision appears impossible.

TABLE 1

VALUES ASSIGNED IN ILLUSTRATION OF PENETRATION OF MINE FIELD

Outcome	Military-System Values	Self-System Values
a. Tank gets through safely in time to engage enemy	$+a$	0
b. Tank searches too long—gets through safely but is of no help in battle	0	$+a'$
c. Tank gets blown up by mine	$-b$	$-b'$

This situation is illustrated in Figure 10c by the crossed ellipses. One ellipse represents the military group system that has the winning and losing states of the war, the other system represents the self-system in which the extreme goals are represented as death on one side and survival on the other.

V. THE RESOLUTION OF CONFLICT

Three procedures are suggested for the resolution of conflict. Actually the first sug-

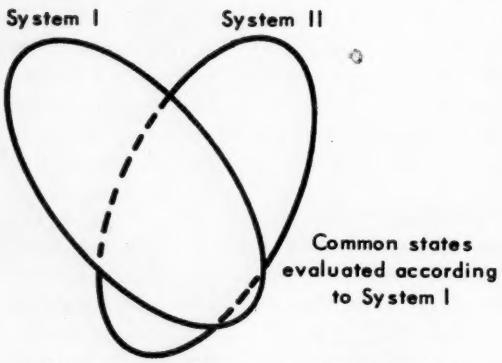


FIG. 12. Dominance and Suppression.

The values of the states shared by the union of Systems I and II are determined by the values of one system only (I) which dominates—the values in the other system (II) being suppressed.

gested method is an extreme special case of the second. These are: "A. Dominance-Suppression," "B. Schism," and "C. Concrescence."

A. Dominance-Suppression

Conflict of values assigned to states that exist simultaneously in two different systems may be resolved by agreeing that the values of one system are to dominate completely over the values of the other system. This situation is illustrated in Figure 12. The dominance of System I may be with respect to all decisions and problems or it may be with respect to certain decision categories only. This situation is a special case of a more general situation to follow and they will be discussed simultaneously.

B. Schism

It may be agreed that one system will dominate only in a portion of the states that exist simultaneously in both systems, that the remainder of this common portion will be dominated by the second system. This situation is illustrated in Figure 13. In effect a boundary is drawn through the disputed union of the systems. A dominance in favor of System I on one side of this boundary, and in favor of System II on the other side results. Should the boundary correspond to the border of either System I or System II a complete dominance-suppression would result.

In general, some type of authority or agreement over or between conflicting systems is required to maintain the schism. This method is a very common means of resolution of conflict, e.g., the whole concept of private property, the jurisdiction of municipal, state, and federal governments. The constitutions of the federal, state, and municipal governments, together with their respective judiciary and police forces, serve to enforce the accepted schisms in our society.

1. Vector values. Underlying schisms in a value structure may exist only for decisions of certain categories. For example, the state and federal governments will have jurisdiction over different types of property, crimes, and civil actions. This is illustrated in Figure

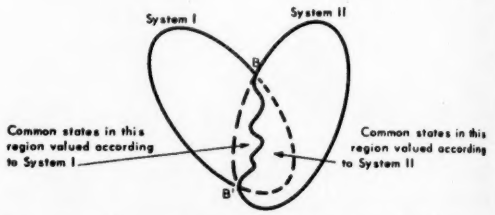

System I
System II
B
B'
Common states in this region valued according to System I
Common states in this region valued according to System II

FIG. 13. Illustration of Resolution of Conflict by Schism.

The boundary of BB' divides the states that are common to both systems into two parts. One part is evaluated according to System I, the second according to System II.

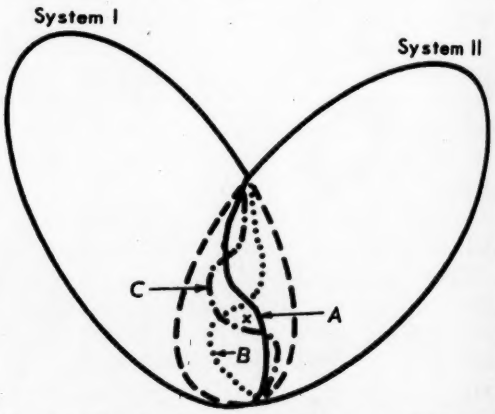


FIG. 14. The Origin of Vector Values by Different Schisms Corresponding to Different Decision Categories.

The boundary A corresponds to one decision category, B a second, C a third, etc. (System I dominates to left of a boundary in that decision category.)

14. The boundary that separates the portion of a common state into two parts such that the values of System I will dominate on the one part and the values of System II will dominate in the remainder is different for the different decision categories. There results for any one particular state in the union of the systems a set of values corresponding to each separate decision category. In Category A the values of state x are determined by System I, in Category B by System II, and in Category C by System II in the illustration. One may confound the situation even further by increasing the number of systems that are involved in the common set of states.

An individual has his self-system, his family system, his community system, his work system, his recreation system, his municipal, federal, and state government system, and his citizenship at large—not to mention such systems as religious groups, social fraternities, professional groups, etc. Complex sets of values are thus built up. The mathematics of multidimensional values will not be discussed here, but can be found in papers by Hausner and Thrall on this subject (24).

C. Concrescence

The third and most important process for the resolution of conflict has been given a name borrowed from Whitehead (29)—"concrescence"—and is illustrated schematically in Figure 15. Conflicts develop in the evaluation of states in the union of Systems I and II. Consider that a third system is constructed enclosing not only the conflicting area but the entire systems that enter into conflict. A new set of value postulates is then set up in such a fashion that a new set of scalar values corresponding to the encircled and inclusive system is established.

The word "concrescence" implies synthesis by encirclement. It also implies the growing together of the value systems. For example, consider nuclear physics in the early 1930's. This may be considered to comprise one system. A second system to consider is the field of national politics. In 1933 these systems were almost entirely separate. With the discovery of nuclear fission and the subsequent invention and development of the atomic weapon, both systems became enlarged and overlapping, i.e., they have concresced. Immediately conflicts develop. Should the nation continue to permit free publication of scientific investigation in the field of nuclear physics? Should nuclear physicists be left free to travel over the world if they wish? Is government any longer independent of the properties of atomic nuclei? Such questions can be resolved by any of the three methods of dominance-suppression, schism, or concrescence.

Concrescence is a creative growth process. Although it has been indicated operationally how concrescence may be considered to occur, there exists no formula, no operational directions serving as a guide, that lead uniquely to a set of new postulates in the supersystem that will allow re-establishment of a scalar value. Concrescence is almost an emotional process and for an explanation of the manner in which concrescence occurs one must look deep into the fundamental psychological structure of the individual personality. It is sufficient to point out that concrescence by restoring scalar values automatically relieves the conflict and its concomitant tensions.

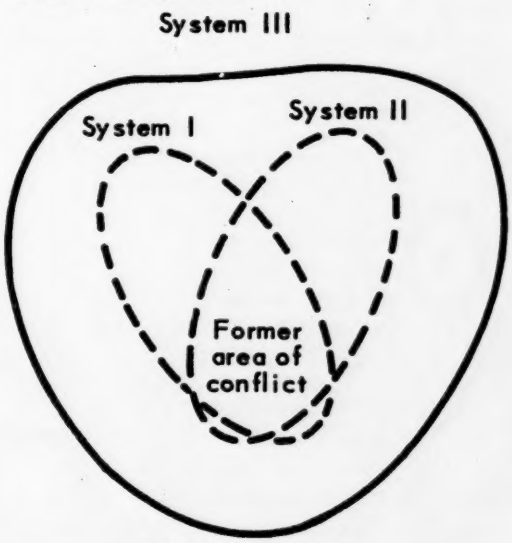


FIG. 15. Concrescence.

A third system enclosing the conflicting systems is constructed; new value postulates are propounded in order to restore scalar values in the total area.

VI. THE DYNAMICS OF CONCRESCENCE

Concrescence is the means of growth and evolution of a value system, which is never the static quantity that may have been implied in the foregoing discussion. In this section some aspects of the dynamics of the growth of value systems as affected by the concrescence processes will be discussed.

A. Growth of a value system

Consider the evolution of a value system. An individual, say, must begin by adopting a minimum set of values—any values. In the language of a behavioral scientist (16) an initial" coding" is assumed. These values are then used on a trial basis for decision purposes. Sooner or later as the experience of the individual increases, the action indicated by these trial values will lead to conflict. It may be a basic conflict between the indicated decisions and fundamental psychological and physical requirements. A conflict produces a stress, which motivates the process of concrescence—either the old value postulates are changed, or new ones added, in such a fashion that the conflict is removed. Thus one characteristic of the evolution of a value system is that it changes in discrete jumps—in a sense that it may be said to be quantized. Occasionally as the value system evolves with increasing experience and increasing numbers of decisions faced, a set of elemental values will be condensed into a single value construct that reduces the number and complexity of the fundamental value postulates. Actually there has been introduced here, even in a theory that claims to be wholly concerned with the form and not the substance of values, two basic assumptions of postulates concerning value systems.

The first of these is that *the resolution of conflict is to be considered an absolute virtue.* The words "inconsistency" and "conflict" imply the same type of tension and will often be used synonymously. It is frequently stated that consistency is a mark of a small mind. This statement would be explained by saying that consistency in a value system in the face of a conflicting situation is to be considered a lack of growth. Re-establishment of over-all consistency by the concrescence growth process is to be considered the highest form of rationality.

In describing the simplification of value systems by the introduction of value constructs, a second basic virtue has been implied; i.e., *the virtue of simplicity.* If two different value systems are equally effective in permitting decisions to be made without conflict the simpler one is to be preferred. In common with usual scientific doctrine we are wielding Occam's Razor.

Although the avoidance of conflict in decision has been described as a fundamental psychological value, it is fairly obvious that this is only one side of a question. In the first place the successful use of values can be accomplished more in a statistical than deterministic sense. Good decisions do not necessarily always lead to the best actions. One must be satisfied with less than perfection—being pleased with a high percentage of successful decisions. Hence any particular value structure must be able to withstand a certain amount of stress and trial and error before any part of it is discarded. The statistical nature of value postulates is acknowledged in the old saying "honesty is the best policy" (not "honesty always pays off"). Individuals learn to adopt a degree of rigidity with respect to value changes. One does not change his fundamental values as a result of the first tension that develops. A stress caused by the conflicting situation must strongly motivate the individual in order to overcome the inertia of rigidity to value change. There seems to be this fundamental built-in conflict that effects the evolution of value systems—the urge to change or adjust values to remove conflict, as opposed to the urge to remain fixed with respect to a value system that has proven successful in the past. Concrescence has been described as a creative and almost emotional process. It is now described as a sometimes painful process. The more a value system has been used successfully in reaching decisions, the greater is the degree of confidence in its workability. This is particularly true if it has successfully withstood conflict stresses in the past without requiring revision by concrescence. Hence one sees a fundamental basic and unavoidable conflict in the dynamics of the

concrescence process itself. Rigidity or pliability may thus become acquired personality traits.

What is the nature of concrescence? How is it accomplished? It is conjectured here that value systems are continually being tested by an imaginative random process. New values are interjected repeatedly on a random trial and error basis (free association?). Rate of occurrence of such imaginative tests is proportional to the tension stimulus brought about by the conflicting situation. We might further conjecture that most of these trials are failures and are rejected. They may be occasionally and accidentally of such a nature that they would remove the immediate conflict but would in turn interject new conflicts in the types of experience already undergone.[6] The object of concrescence is not to remove only the immediate conflict and thereby introduce others, but to remove the conflict without introducing others should the history of the system occur over again. These random trials of new value postulates in some ways are analogous to mutations in a biological evolutionary system where most of the mutations are destructive in character. Occasionally and rarely, however, a trial concrescence is found that relieves the present conflict and introduces no new conflicts. The individual may become suddenly aware that a new value system has been created. Hence the "creativeness" of the process.

Two different characteristics of the value system may be expected to add to the difficulty with which a new concrescence can occur. One of these will be the age of the system (in terms of the successfully accomplished decisions it has rendered). Since any new concrescence must be consistent not only with the immediate situation but with all situations faced in the past history of the system, then new concrescences become increasingly less probable. A second characteristic that may add to the difficulty of performing a new concrescence is the presence of underlying and deeply imbedded schisms in the system. It is not inconceivable to suppose that a complex interlacing of systems deep within a value structure may eventually cause the system to reach a point where further concrescence is no longer possible; i.e., where further concrescence can be reached only through repudiation of the entire value structure, starting anew with a fresh "coding." Such violent and catastrophic behavior of value systems in social cultures can be observed in revolutionary events.

[6] Although it is conjectured here that concrescence is a trial and error process, there exists the possibility that systematic procedures may be set up which will expedite the concrescent process.

B. Conjecture on "Law of Diminishing Weight"

The preceding section has described how *confidence* is built up in a value system. In particular it is suggested that a measure of this confidence is to be found in the number of cases to which a value system has been successfully employed in decision problems without leading to a situation of conflict. As value systems evolve, some value constructs will be older than others. In the first place there will be the originally "coded" values. These are enlarged upon, and occasional concrescences will occur that add new value postulates or that condense a set of elemental value postulates into value constructs. Consider a step in the evolution of such a system. A series of decisions are made on the basis of a value system until a situation of conflict is reached. Concrescence is needed to readjust the value system. Which one of the many value postulates shall be adjusted first in order to remove the conflict? Certainly it will be that value construct in which the least confidence is placed. Since it has been proposed that confidence exists in proportion to the number of times of usage of the value construct, the first construct to be suspected in approaching a new concrescence is that one which has been subjected to the least usage. In frequent cases this will be the newest construct in the system. Thus, as the value system is reviewed for possible alteration, it is not the long-established values that are adjusted, but those that were established most recently or that for some reason have not been subject to use. This practice will give an overweighting influence on the course of

evolution of the value system to the oldest values. Hence, the evolution eventually takes up the form that the old values are seldom tested, only the newest ones. Thus the very first values postulated for the system have an overwhelming weight in the course of evolution of the value system. The newer constructs have an effect on the evolution of the value system that diminishes with their degree of newness. The originally "coded" values in particular will have the greatest weight. This effect may be called a "Law of Diminishing Weight."

1. Reality. Are value constructs real or merely imaginative constructs? Are they arbitrary or absolute? This question will be more fully discussed in the section on the "Epistemological Problem in Value Theory." It is, however, natural to introduce here a concept of *ordered reality. Constructs* shall be considered *real in proportion to the confidence placed in them*. Thus a measure of the reality of a value construct is the number of times it has been successfully (without conflict) employed in decision purposes.

C. Concrescence and the scientific method

In this section the logical congruity between the evolution by concrescence of value systems and the so-called scientific method will be discussed. Consider first the scientific method (15) in building up a scientific theory. In Figure 16 a schematic representation of the scientific method is illustrated.

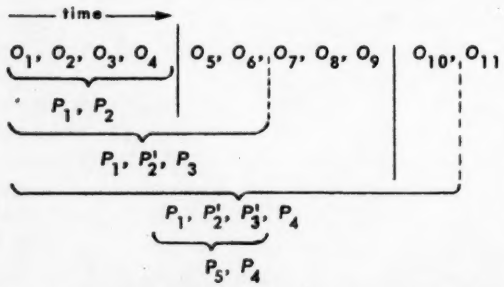


FIG. 16. A Schematic Representation of the Scientific Method.

O_1, O_2, etc. are observations; P_1, P_2, etc. are postulated constructs. The vertical lines represent inconsistencies introduced by new observations, which require revisions of the scientific postulates. P_5 represents a construct of higher abstraction, or order.

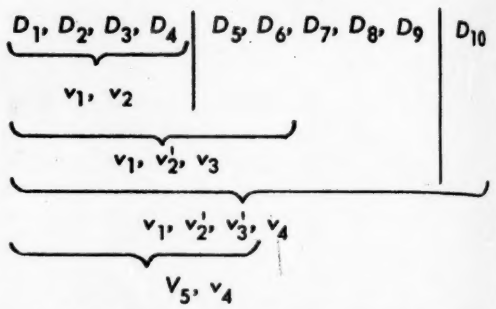


FIG. 17. A Schematic Representation of the Evolution of a Value System by Concrescence.

D_1, D_2, etc. are decision problems. The v_1, v_2, etc. are the value postulates consistent with the decisions. The vertical lines are conflicts (inconsistencies) appearing when a new decision is faced. A new value system is reached by concrescence; the newest value usually being the one adjusted. Occasionally several elemental values are consolidated into one value (moral) construct, (V_5).

A series of observations is first made as indicated from O_1 through O_4. On the basis of these observations, certain postulates are made concerning the nature of the physical world, P_1, P_2.

Sooner or later a new observation, indicated as O_5, is made that gives results inconsistent with the preceding postulates. Thus it may be not only necessary to readjust some of the previous postulates, but to add new ones in order to restore consistency in an ever-widening area of observation, P_1, P_2', P_3. The set of postulates together with the rules for interactions between them constitute the theory. Represented in Figure 16, observations O_7, O_8, O_9, are predicted by the theory. Again, sooner or later, new observations, as indicated by O_{10}, require a new theory or revision of the old theory to restore consistency. Occasionally the theory itself gets so cumbersome with respect to the numbers of postulates and the complexity of relations between them that more general theories or postulates are sought that reduce the number of fundamental notions in the theory, and thus simplify it; e.g. replacement of P_1, P_2', P_3', P_4 by P_5, P_4 in Figure 16.

The evolution of a value system occurs in a similar manner (see Figure 17). A set of

decisions are reached based on a primitive set of values. Sooner or later, as indicated in the figure, a new decision is faced that introduces conflict in the decision. By the concrescence process either new values are added or old values are altered, or both, until consistency is restored and the conflict is removed. In this way one can proceed toward new decisions, with a new set of values. As the history of the system develops, new decisions are faced for which the value system is inadequate, and the value system must be revised by concrescence to remove the conflict.

It is generally cumbersome to resort entirely to elemental values. (Elemental values might be considered as those having to do with fundamental physical and psychological requirements such as survival, sex, etc.) Occasionally a set of these elemental values is condensed into a single value postulate and becomes a moral construct. The process described in Figure 17 may be considered a posteriori as well as a priori. An investigator may review a series of past decisions made in a system and attempt to reconstruct what he believes to be values guiding the decisions that were made. The logical conformity of his methodology with the scientific process is so strong as to lead one to believe the two are identical.

D. Moral and material constructs

It has been described how higher-ordered moral and material constructs can be generated in the course of evolution of value systems and scientific theories. It is proposed here that there is no logical difference between moral constructs and materialistic constructs. One regards the notion of extension in space as reality. Yet, this is a construct that has no qualitative difference from any other construct. For one thing, it is almost impossible to formulate a single thought or take a single action that does not involve this basic construct. Most of our thoughts and actions are consistent with the nature of the space construct. Yet the introduction of the general theory of relativity has shown that a concept having as high a degree of reality as a concept of space needs alteration and adjustment in the light of advanced scientific observation. Space is not the simple Euclidean space normally visualized. It has curvature and local turbulence. Moral constructs differ from materialistic constructs in no observable manner. It is true that some moral constructs have low degrees of reality, but it is also true that some materialistic constructs have low degrees of reality. Consider, for example, the concept of the neutrino in physics. This particle is postulated primarily to keep valid the law of conservation of momentum and the conservation of parity in the spontaneous beta-disintegration of some unstable nuclei.

The process of concrescence has long been recognized. Before Whitehead, it has been discussed by Hegel as the "dialectic," and before Hegel, Plato discussed a synthesis of (a conflict between) the thesis and the antithesis. The word dialectic has been avoided here because of its misuse by the Marxist. The dialectic of Marxism is, in fact, not a dialectic in the Hegelian sense, but the dominance of one class and the suppression of another. True, the conflict is removed, but not by concrescence. If ever the arguments of the materialist are to be refuted, it must be in the realm of metaphysics and metatheory and not on theological or moral grounds, since by the nature of materialism these latter arguments have no significance. It is suggested that value theory provides such a metaphysical ground.

REFERENCES

1. Arley, N. *On the theory of stochastic processes and their application to the theory of cosmic radiation*. New York: Wiley, 1943.
2. Blackwell, D. H. Game theory and the theory of value. First Natl. Meeting, Oper. Res. Soc. Amer., 17 Nov. 1952.
3. Blackwell, D. H., & Girschick, M. A. *The theory of games and statistical decisions*. New York: Wiley, 1954.
4. Bochner, S. *Harmonic analysis and the theory of probability*. Berkeley and Los Angeles, Calif.: University of California Press, 1955.
5. Brooks, F. C. Examples of value in attrition systems. First Natl. Meeting, Oper. Res. Soc. Amer., 17 Nov. 1952.
6. Churchman, C. W., & Ackoff, R. W., with comments by N. M. Smith. An approximate measure of value. *J. Oper. Res. Soc. Amer.*, 1954, 2, 172–187.
7. Doob, J. L. *Stochastic processes*. New York: Wiley, 1953.

8. Feller, W. *An introduction to probability and its application*. New York: Wiley, 1950.
9. Hitch, C. Sub-optimization in operations problems. *J. Oper. Res. Soc. Amer.*, 1953, 1, 87–99.
10. Keyserling, H. v. *Problems of personal life*. London: Jonathan Cape. 1933.
11. Keyserling, H. v. *From suffering to fulfillment*. London: Selwyn & Blount, 1938.
12. Kolmogorov, A. Uber die analytichen Methoden in der Wahrscheinlichkeitrechnung. *Math. Ann.*, 1931, 104, 415–458.
13. Kolmogorov, A. Grundbegriffe der Wahrscheinlichkeitrechnung. *Erg. Mat.*, 1933, 2, No. 3.
14. Lanchester, F. W. *Aircraft and warfare*. London: Constable & Co., 1916, Ch. 4.
15. Margenau, H. *The nature of physical reality: A philosophy of modern physics*. New York: McGraw-Hill, 1950.
16. Miller, J. G. Toward a general theory for the behavioral sciences. *Amer. Psychologist*, 1955, 10, 513–531.
17. Shannon, C., & Weaver, W. *Mathematical theory of communication*. Urbana, Ill.: University of Illinois Press, 1949.
18. Smith, N. M., Jr. Stochastic processes and the theory of value. First Natl. Meeting, Oper. Res. Soc. Amer., 17 Nov. 1952.
19. Smith, N. M., Jr., Value and the decision process. Third Annual Fall Meeting, Oper. Res. Soc. Amer., 19 Nov. 1954.
20. Smith, N. M., Jr., A rationale for operational gaming. Eighth Natl. Meeting, Oper. Res. Soc. Amer., 9 Jan., 1956.
21. Smith, N. M., Brooks, F. C., & Payne, W. S. Application of the theory of value to a study of intrinsic values and cost allocation in the cost analysis of a weapons system. Third Natl. Meeting, Oper. Res. Soc. Amer., 23 Nov. 1953.
22. Smith, N. M., Jr., Walters, S. S., Brooks, F. C., & Blackwell, D. H. The theory of value and the science of decision—a summary. *J. Oper. Res. Soc. Amer.*, 1953, 1, 103–111.
23. Smith, N. M., Walters, S. S. Brooks, F. C., & Blackwell, D. H. Unpublished memorandum, Operations Research Office.
24. Thrall, R. M., Coombs, C. H., & Davis, R. L. *Decision processes*. New York; Wiley, 1954.
25. von Neumann, J. *Mathematische Grundlagen der Quantenmechanik*. Berlin: Verlag Julius Springer, 1932.
26. von Neumann, J., & Morgenstern, O. *Theory of games and economic behavior*. Princeton, N. J.: Princeton University Press, 1947.
27. Walters, S. S. Some theories in the theory of value. First Natl. Meeting, Oper. Res. Soc. Amer., 17 Nov. 1952.
28. Watanabe, S. Symmetry of physical laws. Part I. Symmetry in space-time and balance theories. *Rev. mod. Phys.*, 1955, 27, 26–39; Part II. Q-number theory of space-time inversion and charges conjugation. *Rev. mod. Phys.*, 1955, 27, 40–76; Part III. Prediction and retrodictions. *Rev. mod. Phys.*, 1955, 27, 179–186.
29. Whitehead, A. N. Process and reality; an essay in cosmology. New York: Macmillan, 1929.

(Manuscript received February 16, 1956.)

In reality the scientific method, of which so much is spoken for both good and ill, is whatever means may be appropriate for solving problems in the natural environment. The working scientist brings to bear upon these problems everything at his command —previous knowledge, intuition, trial and error, imagination, formal logic and mathematics—and these may appear in almost any order in the course of working through a problem. The nub of the matter is that the problem be solved. One may go back afterward to analyze what has occurred, and then may generalize it and put it in the form of a logical sequence. But that is not how the thing actually happens; . . . Nothing could be more stultifying, and, perhaps more important, nothing is further from the procedure of the scientist than a rigorous tabular progression through the supposed "steps" of the scientific method. . . .

—The Harvard Report on *General Education in a Free Society*

A CRITIQUE OF KENNETH MARK COLBY'S *ENERGY AND STRUCTURE IN PSYCHOANALYSIS*

by George E. Ruff

Mental Health Research Institute, University of Michigan

Psychoanalytic theory is often criticized because its concept of energy conflicts with that of other sciences. Colby accepts the basic postulates of psychoanalysis, but prefers to redefine them in terms of a modern view of energy. To clarify the relationship between energy and structure, he offers an alternate model of psychic structure.

Energy and Structure in Psychoanalysis[1] is a unique blend of imagination and realism. Beginning with an outline of the pitfalls which await the theorist, Dr. Colby proceeds to offer nothing less than a reformulation of the basic constructs of psychoanalysis. For he feels that traditional concepts of energy and structure encumber psychoanalytic metapsychology with various "logical inelegances." These can be stripped away only by employing new hypotheses regarding the nature of both psychic energy and psychic structure.

As the author points out, present-day psychoanalytic theory links its construct of energy to that of instinct. "Instinctual energy" is described as entering a psychic apparatus, and undergoing various changes. Such a concept has led to a hydraulic metaphor, in which the psychic apparatus, abbreviated PA, is thought of as a series of pipes through which energy flows like a fluid. Colby insists that viewing the PA as a mere conduit system for transporting energy from one place to another is a primitive notion. Instead, he conceives of it as an "active, organizing, transforming, constructing, integrating agency with an energy supply independent of entering stimuli. Incoming stimuli only trigger off processes with the PA".

Colby prefers to adopt the energy model of modern neurophysiology. In contrast to Freud's hydrodynamic energy coneept, he speaks of transformations of small sums of neutral energy which can be measured as continuous electrical oscillations. This raises a problem. Freud, who spoke of various types of energy with different "aims," could use a relatively simple construct of structure. But once energy is looked upon as neutral, the tripartite id-ego-superego model can no longer explain differences in behavior. Colby must solve the problem by introducing a more complex structural model. When this is done, varying behavioral phenomena can be explained as effects of energy transformations in different structural elements of the psychic apparatus.

Before showing how he proposes to accomplish this, Colby is careful to make his philosophical position clear. First he outlines his concept of how living organisms behave. Behavior, he then points out, may be studied on different levels of integration. At each of these, properties appear which are not necessarily explainable in terms of higher or lower levels. Since there is continuity from one level to another, constructs of a higher level may often be converted into those of a lower level. Such a procedure does not then cancel out a level, but rather aids our understanding of the correlations between levels.

In humans, certain functions are best approached on what the author calls a psychic level of conceptual abstraction. They may be

[1] *Energy and Structure in Psychoanalysis* by Kenneth Mark Colby. New York: Ronald, 1955. Pp. ix + 154. $4.50.

viewed as the product of a hypothetical psychic apparatus—an imaginary organization, a construct which aids our understanding of observed phenomena. Reduced to the next lower level, the psychic apparatus corresponds to the functional organization of the central nervous system. But there is never a point-to-point correspondence between the PA and the brain. The hypothetical constructs of the psychoanalyst are on a different level of conceptual abstraction from those of the neurophysiologist, and must not be confused with them.

Colby also points out that the diversity and complexity of the data with which psychoanalysis deals make several sets of concepts necessary for their understanding. Thus, a particular psychological event may be explained in terms of genetic-dynamic, topographic, or economic constructs. From the beginning, ideas of dynamism, structure, and energy have been fundamental postulates of psychoanalytic theory. These form its metapsychology which, Colby is confident, represents an explanation of the essentially orderly nature of psychic life.

ENERGY AND INSTINCTUAL DRIVE

The author states, "dynamic psychology must conceive of psychic activity as the product of forces, and forces involve energy sums. For this reason alone, it is quite necessary that metapsychology have some kind of energy theory . . . [which] aids us in descriptions of change, of activity, of motion, of acceleration and deceleration of processes, of excitement and abeyance." The development of Freud's concept of psychic energy is traced, and it is shown how his combination of the electrical and neurophysiological theories of his time gave rise to the idea of excitation or energy flowing from a sensory to a motor end of a psychic apparatus.

Colby objects to this because it is contrary to the behavior of forms of energy whose effects, unlike those of psychic energy, can be measured. He says, "If our postulated psychic energy represents, at a higher level of abstraction, the energy utilized and released by neurones, then our concept should have some bridging link with physical concepts of energy. Or if the interrelationships between psychic and other forms of energy cannot be adequately stated, at least our views about psychic energy should not sharply conflict with what is known about other forms." Colby makes it clear that psychic energy cannot be converted into other known forms, such as mechanical, thermal, electrical, and chemical energy, and distinguishes it from them by use of the term "cathexis energy," abbreviated CE.

Unlike previous theorists, the author separates the constructs of energy from those of instinct or drive. Freud viewed drive as a force or excitatory stimulus arising from bodily processes. This was said to impinge on the PA and stimulate the organism to specific actions. Unfortunately, drive (*Trieb*) was translated into the English "instinct" (in German, *Instinkt*), which has the connotation of an innate, total pattern of behavior. Freud seldom used it in that sense, but defined it simply as a construct whose "... source is a state of excitation within the body and [whose] aim is to remove that excitation." He then spoke of anatomical sources of drives, and at first named these drives according to their sources. Thus, sexual drives were felt to arise from somatic erotogenic zones and self-preservative drives from the cells of the "nutritional system."

Later, Freud reclassified drives according to descriptions of the behavior they evoke—either sexual or aggressive. Some analysts have followed the first division and others, the second. But almost all use the words "instinct" or "drive," more or less interchangeably, to indicate a force which arises from the body, impinges on the PA, and thus impels the organism to specific actions. Most analysts also follow Freud in considering drive "a certain sum of energy forcing its way in a certain direction." Drives are then classified according to their aims, rather than their sources or objects.

Although Colby uses the terms "instinct" or "drive" in the same sense as Freud, he feels that the connection of energy with instinct "brings difficulties into formulating a logically consistent energy theory, since all other forms of pure energy, as we know them, have no aims. Energy alone simply dissipates." He thus prefers to re-define the

constructs of energy and instinct. To him, cathexis energy is neutral, like all other forms. Only when it operates in some kind of structure can it produce effects.

Drives are consequently defined not as pure energy sums, as traditionally formulated in psychoanalysis, but as structural components of the psychic apparatus activated by cathexis energy. More specifically, drives refer functionally to the active influence of certain microunits of the psychic apparatus. These units, called "drive schemas," are invested with cathexis energy of their own, which is influenced by incoming physiological stimuli, but is not dependent on them. Colby notes that this formulation of drives contrasts sharply with Freud's view of a chaotic and haphazard id.

While drives have been variously classified according to source, object, and aim, Colby prefers the latter. He states, "The aim of a drive is to regulate, within their optimal ranges, oscillations which, by threatening their normal upper or lower limiting values, would bring the organism into a critical state." When activated, drive schemas excite the psychic apparatus to carry out specific actions toward specific environmental objects, resulting in stabilization within these optimal ranges. Thus, drives are viewed as tensions arising from psychological processes. Their aim is reduction of tension, which is then experienced as gratification. But whereas Freud looked upon drives as different types of energies, which lead to behavior changes in the organism by flowing from place to place in the PA, Colby views them as structural elements which produce effects when activated by neutral energy.

These behavioral acts are divided into preparatory and consummatory acts, where the former roughly refer to means and the latter, to ends. Specificity is given by the particular structural units activated. This determines the category of the drive, which is labeled by a verb descriptive of its consummatory act. The two general drive categories are Maintenance (M) drive schemas, and Reproductive (R) drive schemas. M-drives excite consummatory acts which an organism must perform in order to live. They are classified in four divisions: (*a*) to breathe, (*b*) to sleep, (*c*) to ingest, (*d*) to excrete. R-drives, on the other hand, are more complex and plastic. Unlike M-drives, which they overlap to some extent, R-drives may be deferred or only partly fulfilled.

Reproductive drives are divided into (*a*) a drive aim to mate and (*b*) a drive aim to rear. Colby applies the terms "sexual" and "erotic" to all processes of the mating drive, in contrast to Freud's definition of "sexual" as pertaining to ". . . obtaining pleasure from zones of the body—a function which is subsequently brought into the service of reproduction." Freud's definition, however, includes all maintenance activities, such as eating, under the term "sexual," and excludes consideration of the quality of a pleasure.

For this reason, Colby prefers the standard definition of pleasure as a "quality of consciousness, a subjectively-experienced affect-feeling which is recognized as gratifying. . . ." The quality of pleasure accompanying the mating drive is "sexual," and oral and other activities are considered sexual only when they are preparatory actions for this drive. It is often difficult to recognize the sexual nature of such activities in children, where consummatory actions are undeveloped and preparatory actions are used as ends rather than means. Final adult aims and objects are potentials which develop only with maturation. Preparatory actions of drives, however, are given adjectives which relate to eventual consummation, regardless of how distantly they are separated in time.

CONSTRUCTS OF STRUCTURE

Colby proposes a new structural model of the psychic apparatus. He points out that Freud offered three different models of the PA, using words and rough diagrams. The first of these, essentially a neuronal model, appeared in *Project for a Scientific Psychology* (1895). The second was described in *The Interpretation of Dreams* (1900), where the psychic apparatus was compared to an optical system, containing some neuronal elements. This involved a dynamic organization, for cathexis energy flow could take place between a series of successive systems.

Freud's third model was the familiar one, utilizing the constructs of id, ego, and superego. "Conscious," "preconscious," and "unconscious" now became qualities in a descriptive sense rather than systems occupying fixed spatiotemporal positions, as in the second model. He hoped these constructs would be "modified, corrected, and more precisely determined as more experience is accumulated and sifted." Nevertheless, no new models have appeared, most present-day theoreticians having preferred to clarify and reformulate the tripartite constructs.

This task is approached realistically. The author is always aware that models are designed to show relationships between the parts of a whole. He also warns that models are likely to be mistaken for theory, "whereas they act only to illustrate a theory or certain of its aspects," and adds, "Being imaginary, our models are not necessarily true or false but only apt or inept." He quotes a pertinent comment by Freud: "We are justified, in my view, in giving free rein to our speculations so long as we retain the coolness of our judgment and do not mistake the scaffolding for the building."

Colby freely acknowledges that "one model can explain certain functions better than another, but fail in some respects to equal the assets of a third, which is known to be inferior to the second in other respects, and so on." This is true, he feels, because psychoanalysis has accumulated such a mass of observations on human experience that one model or set of constructs cannot hope to explain it all.

The functions of a good model are then described. It should "indicate those variables we consider of primary importance and those which are secondary or derivative. Next it should illustrate how the PA integrates . . . internal biological and external environmental [conditions]. A third requirement for a model is that it allows for a suitable account of maturation and experiential development. And while establishing universals which hold for everyone everywhere, a structural model must provide for individual uniqueness and societal variations. . . . Lastly, a structural model should be suitable to indicate how the PA functions. A model of structure is inert and uninteresting unless one also specifies how it works. Similarly, models of function or CE distribution remain tenuous until supplemented by structural constructs which state what is doing the functioning."

With this statement of his goal, Colby presents his cyclic-circular model. This is a complex structure which can be clearly grasped only by reading its complete description. For purposes of discussion, however, a brief outline is given here, supplemented by one of the diagrams found in the book. The model consists of ten large systems, which are made up of microunits called "schemas." The schemas form a sort of reticulum or matrix, containing concept-meanings in codified form. Each schema has a characteristic frequency of pulsation—its cathexis energy period. By modifications in their periods, schemas constantly change in design.

Messages move through the psychic apparatus in much the same way a patterned figure crosses an electric sign by successive lighting and dimming of bulbs. When signals enter the psychic apparatus, "cycle patterns," which have characteristic CE periods of their own, are set up to bear messages. A chain reaction of CE processes then fires off, moving cycle patterns through the series of systems. Each schema in the transport systems momentarily adopts the CE period of a cycle pattern as its own, and quickly passes it on unchanged to the next schema. In storage systems, however, the CE patterns of schemas interact with that of the moving cycle pattern, and facilitate or inhibit its concept-meaning content. If the CE period of a cycle pattern is synchronized, facilitation occurs. If it is desynchronized, inhibition occurs. Unlike transport schemas, storage schema CE periods may in turn be modified by cycle patterns. Thus, results of past experience are stored and can subsequently influence the response to new incoming stimuli.

The model can be better understood after following the course of a cycle pattern through the psychic apparatus. The diagram in Figure 1 will help illustrate this path. It should be noted that systems named with

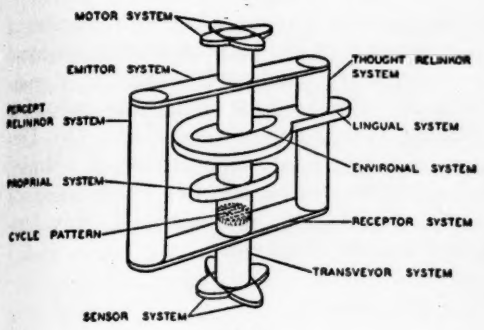


FIG. 1.

suffixes of *-or* have transport functions, while *-al* suffixes indicate storage functions.

The sensor system receives physical energy sums from drive sources in both external sensory organs and internal physiological processes. These signals excite sensor system schemas, representing particular bodily areas or processes. A cycle pattern is now set up and moves along the transveyor system to the receptor system, where it is joined by messages from previous cycles. From here it goes to the proprial system, where the first storage schemas are met. This system contains intrapsychic representations of the self.

Drive source messages from the cycle pattern may excite drive schemas, which represent innate drive actions. If similar messages in previous cycles produced actions which proved gratifying, the drive schemas will have been rearranged in such a way that the incoming cycle pattern will be facilitated. But if previous actions induced discomfort, the cycle pattern is inhibited to varying degrees.

The modified pattern next enters the environal system, where structural representations of the external environment are found. Again facilitory and inhibitory transformations occur. Either drive object schemas will excite action to obtain required objects from the environment, or alarm schemas will lead to avoidance of objects or conditions considered dangerous. The cycle pattern is then carried through the transveyor system once more.

In the emittor system, it splits into three segments. One segment, which contains motor messages, continues to the motor system. Schemas representing specific bodily motor areas are excited by action messages, and signals are sent out to activate motor organs. The other segments enter the percept and thought relinkor systems, where consciousness appears for the first time. The percept relinkor system allows awareness of what is perceived in the environment, while the thought relinkor system allows awareness of thoughts. Cycle patterns in the thought relinkor system are further modified by passage through the lingual system, whose schemas contain concept-meanings relating to words and rules of language usage. Messages are now perceived as linguistic thought, whereas previously they were conscious but nonlinguistic.

The cycle pattern is thus subjected to the modifying influences of proprial, environal, and lingual systems, and in turn, modifies the schemas with which it interacts. After leaving the relinkor systems, the reconstructed cycle pattern is returned to the receptor system for rerunning. By a feedback mechanism of this type, the psychic apparatus can respond to changes within itself. Furthermore, the actions produced by signals to motor organs give rise to signals, which enter the next cycle via the sensor system. It can thus be seen that events in the environment, within the organism, and within the PA itself, all contribute messages to a cycle. In turn, the PA sends out signals to the environment and organism, and reruns others. The term "cyclic-circular," therefore, is appropriately applied to Colby's model. He feels that "as Freud's models reflected the scientific climate of his time, so the cyclic model inescapably represents the ideas and imagery of our time."

ENERGY, STRUCTURE, AND FUNCTION

Colby points out that a metapsychology must first attempt to "reduce observed wholes into parts simpler to understand. After such a hypothetical reduction is accomplished, the next step . . . is to formulate an equally hypothetical reintegration, showing how the parts combine and work together to produce an observed whole." Consequently, he takes up examples of observable psychic properties and indicates, in the

language of the model, how they represent the interactions of cathexis energy and structural elements containing various concept-meanings. In doing this, he deals with percept, thought, action, affect, reality testing, and symbolization.

The details of this presentation cannot be given here except for a few fragments to illustrate Colby's approach. For instance, development of the reality principle is explained as follows: "With maturing and development of schema-level complexities," certain action messages can be suspended, while percept and thought messages are rerun and alternatives considered. "Action is delayed until probable responses of self and environment are sorted out and a safe, as well as gratifying, moment for action can be selected."

With the aid of the model, Colby differentiates five components of the processes involved in affect. Underlying these is the idea that pleasurable affects result when CE periods, synchronized by the proprial and environal systems, become conscious in the relinkor systems. When the CE periods of thought and percept are desynchronized, one consciously experiences unpleasure and dissatisfaction. Synchronization is thus a go-ahead signal experienced later in consciousness as pleasure. Unpleasant affects, resulting from dysynchronous CE periods, "indicate that action, thought, and even percept should be avoided."

The phenomenon of stratification of affect is particularly well handled by the cyclic model. For instance, a patient feels anxious, develops shame over his fearfulness, becomes angry for having been ashamed, feels guilty because of his anger, and finally becomes anxious again over his guilt. This produces an "affect spiral," accounted for by successive runs and reruns of processes through the psychic apparatus.

THE FREUDIAN AND CYCLIC-CIRCULAR MODELS

Colby warns against applying the model to all psychoanalytic theory, stating that it is "only for the logical and metapsychological interrelating of basic postulates of psychic energy and structure." However, an inkling of the cyclic model's explanatory power is given by even a brief comparison of it with some of the constructs of traditional psychoanalytic metapsychology. For convenience, this is done under the following headings: (*a*) The Unconscious; (*b*) Dynamics—including energy, instinct, libido; (*c*) Genetics—including the importance of early experience, psychosexual development; and (*d*) Structure.

The unconscious

Like Freud, Colby describes unconsciousness as a quality of thoughts and percepts. In the cyclic model if past messages have been followed by unpleasant results, subsequent similar cycle patterns will be desynchronized. Once filtered out by inhibitory conditions in the proprial and environal systems, cycle patterns fail to reach the relinkor systems, and simply never come to consciousness. The preconscious, it may be added, consists, at any given moment, of cycle patterns which have not yet reached the relinkor systems.

Dynamics

Before comparing Colby's concept of energy with that of traditional psychoanalytic theory, the development of the Freudian energy construct must be considered once more. From the beginning, Freud insisted that a complete explanation of psychological phenomena had to include economic or quantitative aspects of these phenomena. He defined the economic standpoint as "one from which we try to follow out the fate of given volumes of excitation and to achieve, at least relatively, some assessment of it." Freud thus thought in terms of changes in an axiomatic "quantity," which must underlie behavioral changes observed empirically.

Coupling this conviction with concepts of electromagnetism of his time, he wrote in *Project for a Scientific Psychology*, "... among the psychic functions there is something which should be differentiated (an amount of affect, a sum of excitation), something having all the attributes of a quantity—although we possess no means of measuring it—a something which is capable of increase, decrease, displacement and dis-

charge, and which extends itself over the memory traces of an idea like an electric charge over the surface of the body. We can apply this hypothesis, which by the way already underlies our theory of 'abreaction,' in the same sense as the physicist employs the conception of a fluid electric current."

As long as such a "quantity" is defined only as an "affect" or an "excitation," there is no conflict with modern concepts of physical energy. It is perfectly permissible to speak of the flow of affect or excitation from place to place in a hypothetical psychic apparatus. We may also mention that for a quantity of something to flow from one point to another, there must be a difference in pressure between the two points. If we consider the quantity of affect or excitation analogous to an electric charge, we may say that it flows because of a potential difference in our apparatus. We may go further and postulate that expenditures of energy are necessary to produce this difference of potential. But in an electrical system, the charge which flows is not synonymous with the energy transformation which produces the flow. Similarly, in our psychic apparatus, the affect or excitation which flows must be differentiated from the energy which is expended to make the flow possible.

This becomes clear when we consider some elementary concepts of physical energy. Energy can be viewed as a "quantity" which has the capacity to produce a change in the quantity of something else. In order for a quantity of something to change—either in magnitude or in direction—there must be a change in the level of energy. But this is not to say that energy can be both the quantity capable of producing a change and the "something which is changed."

Unfortunately, this is exactly what Freud did in defining libido as the energy arising from excitation in the erotogenic zones of the body. For libido was also the substance which flowed from place to place within the psychic apparatus, with alternate binding and unbinding to specific objects. Freud hoped that ultimately these changes would be measured. Libido, however, cannot be both the quantity whose changes are measurable and the energy with the capacity to effect such changes. If it had been defined merely as a form of energy produced by sexual processes, libido could be considered a capacity for causing a change in something else—excitation or affect, for instance. If, on the other hand, libido is to be the quantity whose variations and flow are measured, something else must provide the energy.

Colby, however, does not raise this problem of the confusion resulting from interchangeable use of terms like "libido," "excitation," "affect," "quantity," and "energy." His primary objection is that the Freudian model couples instinct and energy, which then flow in particular directions. This is also contrary to what we know of other forms of energy, for energy in the physical sciences is a scalar, not a vector. In Colby's model, therefore, cathexis energy is neutral. It becomes distributed throughout all the microunits of all the systems, and no constant level of energy is assumed. This contrasts with Freud's tripartite model, where energy is distributed like a head of water—a falling level in one part of the container must be accompanied by a rise in another. Changes in energy in the cyclic-circular model result from increases or decreases in the frequency of cathexis energy pulsations.

The CE activity of Colby's schemas is autonomous within the psychic apparatus. Energy is modified, not stopped and started by incoming stimuli, and is not dependent on instinct or drive. Viewing energy in this manner is a distinct advance over conventional psychoanalytic metapsychology, where some theorists have found it necessary to postulate undifferentiated, aggressivized, libidinized, deaggressivized, and delibidinized energy. Colby simply divides drives into maintenance and reproductive categories. This represents a break with traditional psychoanalytic theory, which usually speaks of sexual and aggressive drives. With energy dissociated from the concept of instinct, and viewed as neutral, he considers aggression a mode of behavior—a means to an end. It is not an instinct or drive, but a

preparatory type of behavior toward any drive object.

Genetics

The description of the cyclic model applies to the psychic apparatus in its adult form, or at least after the development of language. Colby suggests that growth mainly takes place in the microunits—especially in the proprial and environal systems. The phenomena associated with psychosexual development occur because mating drive schemas have a maturational history extending over many years. Before a final stage is reached, several organizational stages are passed through in sequence, where areas of the body successively gain and lose priority in erotic predominance. Colby's model, therefore, provides a mechanism for explaining the developmental data psychoanalysts have observed. Although Freudian theory stresses the importance of learning and past experience, it does not account for the storage of memories.

Structure

By now it is clear that the structural aspect of the cyclic-circular model is both more complex and more highly organized than in the Freudian model. This reflects Colby's conviction that metapsychology must attempt to explain the orderly nature of psychic life. Since reality is organized, a psychic apparatus which represents an element of reality should also be organized. In addition, if energy is neutral, drive aims can no longer be determined, as in the Freudian model, by differences in instinctual energies. New structural components must be introduced to explain the variations in behavior resulting when the organism acts to reduce the tensions associated with different drives. A more complex structural model, however, is not necessarily disadvantageous. It may actually aid in the formulation of theory. Furthermore, as our knowledge of the real world grows, increasing complexity of structure has been found on all levels of integration, so that Colby again brings his model of the psychic apparatus into closer alignment with what has been learned about the brain.

To carry out this idea, Colby abandons the idea of an id serving as the repository of a chaotic conglomeration of drive energies, aims, and objects. Instead, he distributes drive components throughout all systems of the psychic apparatus in an organized fashion. On the other hand, the ego, in the Freudian model, is not a structural entity in the same sense as is the id, for it summarizes a group of functions, some of which represent the activities of id or superego. An additional problem is that so many ego functions have been described that it loses its value as a basic construct. Rather than lump these into one division, the cyclic model "pictures a sequence of systems, named and defined according to structural content and function. . . ." Colby considers the superego the clearest construct in the tripartite model, and extends it by again postulating differentiated systems with specific concept-meaning content. One of these is found in the environal system, where schemas have facilitory and inhibitory influences "similar to the functions of the superego, which both encourages and prohibits."

Colby points out the tripartite model's advantage in expressing theories of conflict, since it permits a clear visualization of conflict. In the cyclic model, where conflict is represented as a phenomenon of microunits of structure, its pictorial representation is difficult. He suggests that, like other schema-level functions, conflict may eventually best be described in "algebraic or other non-geometric terms." Once again, however, Colby approaches present-day concepts of the function of the brain. For instead of flows of relatively large quantities of energy from place to place in the psychic apparatus, the neurophysiologist speaks of transformations of minute quantities of energy between small structural units. Thus, conflict probably is a "microphenomenon."

EVALUATION

Definitive criticism of Colby's ideas must center on how well they work. Only after the model has been put to the test of stimulating and facilitating theoretical formulations, and eventually of reducing clinical data to

understandable terms, can its effectiveness be judged. One wonders if it will ever receive the trial it deserves. Both theorist and clinician are sure to find that using a new model is like learning a new language—a process resisted by those who think their native tongue serves them perfectly well. Furthermore, during the past thirty years a staggering mass of observational data have been recorded and interpreted in terms of the Freudian model. The translation will not readily be undertaken.

It is evident, however, that Colby's model has at least two important advantages: it clarifies present concepts of energy and drive, and at the same time, eliminates those aspects of traditional metapsychology which conflict with what is known about brain functions on other levels of integration. The value assigned by the reader to the second point will inevitably depend on his feelings about the usefulness of models in general.

Colby has employed a formal model—one which does not directly represent any aspect of reality. Ideally, however, propositions involving relationships between components of the model will be applicable to other systems, where either analogous relationships exist or where there is a possibility that such relationships may exist. When this is the case, we hope that deductions can be made from the model which lead to the discovery of real things. In contrast, Freud's constructs are hortatory models—figurative representations of the psychic apparatus which do not pretend to draw an analogy to something which exists in reality. Consequently, conflicts between his energy model and our knowledge of other forms of energy may seem unimportant. However, such a model lacks heuristic value.

Viewed from this standpoint, the cyclic-circular model is a step forward. Like Freud's, Colby's model is a mechanical apparatus whose functions represent relations that have been observed and ordered in clinical work. But by attempting to make his model compatible with, for example, the neurophysiologist's concept of reverberating circuits, Colby has created a possible opportunity for advance in both psychology and neurophysiology. If the model should prove an apt one, the psychologist might employ analogies to what is known in neurophysiology as leads for research to fill in gaps in our knowledge of psychology. Conversely, if constructs on the psychic level of integration are logically homologous to those of physiology, psychological knowledge may stimulate discoveries related to the workings of the brain.

It is granted, then, that since libido fails to correspond to what is known of the flow of physical energy, it should be eliminated from our model of the psychic apparatus. But it may be argued that Colby's model, which does not deal explicitly with the effects of energy transformations, could dispense with the concept of energy entirely, and still order the phenomena observed. For although he rightly insists that energy can produce effects only by operating in some kind of structure, he does not make it clear just what effects are produced. Frequencies of pulsation in the schemas vary, cycle patterns become synchronized and desynchronized; but is it necessary to say that "cathexis energy" is pulsating, synchronized, or desynchronized?

In fact, as Colby himself points out, it may be disadvantageous to employ the term "energy," since we may then be tempted to speak in terms of properties of physical energy before we have any right to assume analogous properties exist on the psychic level. It is true that Colby stresses the value of an energy construct for aiding us in descriptions of change, motion, acceleration, and the like. But again we may ask, "Changes in what?" Before speaking of energy changes, we should have some concept of the nature of the quantity in which their effects are seen, even though it may not be measurable as yet.

Sixty years have passed since Freud spoke of something "among the psychic functions . . . which should be differentiated (an amount of affect, a sum of excitation), something having all the attributes of a quantity . . . ." This elusive quantity remains undifferentiated. When its role in the psychic apparatus has been defined—when we have discovered some kind of entity that is

altered in magnitude or direction—the next step must be to postulate a psychic energy whose transformations make such changes possible.

As was true with Freud, however, Colby's faith that "quantities" exist has prompted him to use the concept of energy, even though he cannot yet specify how it is essential in the operation of the model. He explains that although we cannot measure quantitative differences in psychic life, clinical observations lead us to believe that they exist, and suggests the "cathexis" as the standard unit of CE, so that varying degrees of activity can be compared. Thus, even though quantities are not measured, intensities proportional to frequencies of pulsation can be specified.

Perhaps Colby's reasoning can be justified in the following manner: Our observation of what appears to be goal-directed behavior, coupled with a conviction that the organism is more strongly motivated to do some things than others, can be said to make necessary some concept of a "psychic difference of potential." If changes in some kind of quantities resulting from this potential difference were known, and could be measured, we would be doing much the same thing as the physicist, who recognizes and measures energy transformations by the effects they produce. Although we have not yet defined such quantities in psychoanalysis, we nevertheless feel they do exist. Not only that, but our knowledge of the energy needs of neurones, whose discharges are the basis of behavior, suggests the need for a psychic analogue. A structural model of the psychic apparatus, then, must include some concept of psychic energy, even though no direct use of it is made at present.

In summary, Colby has sought to clarify the logic and metapsychology of psychic energy and structure. His success has been noteworthy. At the moment, many of the potential clinical applications of his constructs are implicit. Further theoretical advances, formulated in terms of the model, will be necessary before its clinical usefulness becomes explicit. A more immediate application may be found in permitting a greater degree of integration between cybernetic and information theory, on the one hand, and psychoanalytic theory, on the other, than has been possible with the tripartite model.

But of greater importance is the model's potential for linking psychological and physiological levels of integration. Psychological descriptions of behavior should never ignore behavior's dependence on the functions of nerve cells. Colby's insistence that the design of a psychic apparatus must not conflict with knowledge of the brain thus keeps psychoanalysis firmly tied to its biological roots. This was Freud's goal. By bringing the model up-to-date, Colby's work is in the mainstream of psychoanalytic progress. And what could be more appropriate at present than a model based on such ideas as the spread of electrical fields, reverberations of nerve nets and loops, and synchronization or desynchronization of electrical rhythms and discharges?

Furthermore, the breadth of Colby's constructs, particularly with their redefinition of concepts of instinct, sex, and aggression, may bring a variety of analytic viewpoints closer together. For his model does not impose any specific point of view on the theorist who employs it. Followers of Freud, Sullivan, Horney, and others, therefore, may all order their data in terms of the cyclic-circular model, although they may differ in their interpretations of these data.

Colby clearly considers his ideas only an alternative model of the psychic apparatus. He is careful to point out that the advantages of the tripartite model for understanding and communicating ideas about patients may well make it better suited than the cyclic-circular model for clinical work for some time to come. In fact, he suggests that "as in other sciences we may soon find a need in psychoanalysis for two sets of constructs and two languages which represent the differentiation between pure and applied aspects of our studies. Metapsychological hypotheses represent the pure aspect of our science and they are temporally much in advance of practical therapeutic applications. If we progress, there will necessarily be a wider and wider gap between pure metapsychological and applied clinical psy-

choanalytic theory .... New structural formulations must be weighed and debated long before they have bearing on clinical techniques of therapy. To avoid enshrining the few models we already have, we should find use for multiple models in psychoanalytic theory. That one is discarded in favor of another indicates only a scientifically healthy temporariness of those fractions of probability we call constructs. In a picturesque moment, Mach ... stated, 'Theories are like withered leaves which drop off, having enabled the organism of science to breathe for a time.'"

Between you and those whom you would serve through science the same gulf lies as between the priests and those they would benefit. The protest against science is not yet so loud, I grant you, as that against religion, but it is the same in kind, and it is growing. Scientists are as eager to do our thinking for us as ever the Church has been, they are just as ready to use force to make effective the truth as they see it, and they keep their scientific spirit to themselves as effectively as the priests keep their priesthood. They look upon themselves as a caste, and in the name of science they presume to dogmatize outside of their field, exactly as the priests once did. ... We grow rich it appears in the results of science, but poor in its spirit. ... But the criticism is justified in the places where science is avowedly engaged in handing on her torch—in your college, for example, where almost all of you studied the sciences and almost none of you was suspected by anybody of being scientific.

—John Erskine in *The Call to Service*

ROUND TABLE DISCUSSIONS ON BEHAVIOR THEORY

Physicians often publish in their journals verbatim transcriptions of clinical pathological conferences in which they make public the process by which they come to their diagnostic conclusions and carry out other clinical functions. Lawyers often put complete transcriptions of their legal reasoning on the record. With scientists this is not so common. In this department edited segments from round table discussions concerning behavior theory are presented. These discussions are largely informal and unrehearsed, and they illustrate more the activity of the development of such theory and the interplay among individuals and disciplines, rather than finished products. At points their value more likely resides in the liveliness of the interchange than in the accuracy of the conclusions.

Theory Seminar

Participants, all associated with the University of Michigan: Ralph W. Gerard, biology (neurophysiology), Mental Health Research Institute; Donald G. Marquis, psychology, Department of Psychology; James G. Miller, psychiatry and psychology, Mental Health Research Institute; Theodore M. Newcomb, social psychology, Departments of Sociology and Psychology; Anatol Rapoport, mathematical biology, Mental Health Research Institute.

September 21, 1955

Miller: Since this is the first session of this group, we should at the beginning determine the best method of procedure and just where we should start. I'd like to ask Gerard to outline what he means by "boxes." It should be emphasized that in organizing our thoughts about behavior theory we are at the beginning at least, more interested in the forest than the trees.

Gerard: That's most important. I am sure that only by the most intense determination can a small group like this force itself to stay at a patrician level of generalization in order to do the job of systematically looking over the forest.

Now to discuss what we mean by "boxes." Here they are, in Table 1, the intersections of these columns and rows.

It should be most stimulating to discuss "institution," and probably also "species" since they do not seem quite like the other levels. I had originally included equilibrium as a fourth column (and I have listed it above with a question), but equilibrium breaks down nicely into architecture or function.

Miller: Why did you consider equilibrium to be as important as, or comparable to, those other rubrics?

Gerard: The three attributes that would seem important, as you look at the way biology divides up, are dynamic equilibrium (maintenance of homeostatic processes), specific synthesis (building up of substance and pattern in growth and reproduction), and adaptive amplification (which is, of course, behavior and its trigger feedback aspects). So naturally I was inclined to start with those and carry them over. Perhaps action can be divided into conservative or routine action and creative or innovative action.

Newcomb: Surely in talking about history you get into problems of architecture and action. There could be a history of architecture, a history of action, etc.

Gerard: I think not. History means the

path of the system in which you are interested. At the organismic or individual level, this would involve embryology, learning, etc. At the group or species level it is evolution. Action would essentially be those behavioral changes in a system which are not cumulative or secular or irreversible. When a muscle contracts, this is clearly not part of history. When a muscle grows from an embryo to an adult, it is history. When it hypertrophies with exercise it is moving into history.

Our hope is that we can formulate for ourselves some of the critical attributes of each box and the extent to which what is proved for one box will be seen down the line. For instance, when we speak of growth or differentiation of individuals or origin and evolution of species, of learning in the individual or structuring of a group or establishing a new community, or whatever it may be, to what extent are we dealing with the same basic process? Just what is the extent of commonalty and what are the significant differences?

Rapoport: It seems to me that every one of these columns is subject to further breakdown and perhaps this should be done at an early moment. For instance, suppose you wish to define social distance. You look for criteria and measures. Are these measures cardinal or ordinal? In other words, are they actually units or simply rank orderings?

Miller: We need to be careful in considering such notions as *roles* or *stresses*, to distinguish whether they are real entities in space-time or whether they are other sorts of aspects of systems. I hope we can carefully differentiate dimensions of space and time, dimensions of information, and dimensions of mathematical conceptualization, distinguishing among systems of different sorts.

Rapoport: I wonder if there is any such sharp distinction. You say a thing is real if it exists in space-time; in other words, you can give it space coordinates and time coordinates. Can you give it mass coordinates?

Miller: Yes.

Rapoport: Then mass is real. How about electric charge? Is that real? How about

TABLE 1

Level	Architecture or Structure	Action, Function, Dynamics	History	(Equilibrium?)
Molecule				
Cell				
Organ				
Individual				
Group				
Species				
Society or Eco-group				
(Institution)				
Biota				

chemical potential, the affinity of an element to combine with another? Is that real?

Miller: To the extent that it can be operationally determined, yes.

Newcomb: In your September, 1955 *American Psychologist* article, Miller, it had to be measured in CGS (centimeter-gram-second) units or known derivatives to qualify as real in your terminology.

Rapoport: We may not always be able to work on that level. Suppose that we define social distance quite arbitrarily as the number of times one person pays another a compliment or an insult. Both would be a measure of social distance, one positive, one negative. Now the number of times would have dimensions per unit of time, so fitting into the CGS system. But how about insults and compliments? They have to be evaluated semantically.

Miller: Here is where we have to be careful. As an empiricist I am interested in how I can measure. There certainly has to be a set of semantic rules, since everyone is coded differently. We often assume, mistakenly, that everyone is coded the same way because we all speak English or all understand the gesture-affect language. But this is far short of precise science.

Newcomb: Suppose such measurement could become precise, but not in CGS units; then what?

Marquis: Then this is the simple criterion

of objectivity and agreement among observers. I think the question that is put to Miller now, is whether it is objectivity that is the criterion or whether CGS is the criterion.

Gerard: I'll answer for Miller. His faith is that, whether or not one can actually do the reducing into CGS at the moment, CGS type units will emerge as one gets sufficiently objective.

Miller: That's a good answer. Units for measuring temperature existed and were convenient in physics prior to the time that Maxwell and others did the work which made specific the relations between temperature units and CGS units.

Marquis: Let me offer a kind of test case. Is an attitude, let's say a man's attitude toward the Republican Party, real? As you normally think of it, it isn't measurable in CGS, yet it may be operational, and in certain circumstances it may achieve very high precision.

Miller: Actually in my article I used the phrase "*u* units." These embrace not only CGS units but information bits and certain other types of informational units.

Marquis: Isn't what Miller is offering a program for research rather than a rationale for measurements we can make at present?

Miller: That's quite true.

Newcomb: This discussion is reminiscent of a fascinating bit of sociological history that had to do with an attempt in the late 1920's and early 30's to develop a methodology of objective recording of behavior. Dorothy Swaine Thomas worked on this most assiduously for about five years, after which time she concluded that one could have his choice of being very objective about matters that are utterly inconsequential or not achieving any decent objectivity about the matters one cared about.

Gerard: Let me reply to that comment by mentioning an experience of mine with another group similar to this. We started by turning in various propositions about behavior, and then we decided to limit our observation to an institution with which we were all familiar, the university. We started tossing in propositions, such as: the department is the strongest unit of the university (the reason being that one identifies primarily with his professional group more than with his local group and that departments parallel professional organizations). These propositions were turned into mincemeat by such questions as what do you mean "strongest," etc. After several weeks of this, we found we had refined the propositions to the extent that they were accurate but meant little. But, despite this, we found that, without pinning ourselves down to a verbal precision, we could reach an operational understanding that enabled us to predict indicators and devise indices which were quite experimentally valid.

Let us summarize some of the things we seem to agree need attention. First, making the table itself—are these the right boxes? Second, determining the things which can be said about a box in general; perhaps a more precise definition of the unit, the entities of each level, and certainly a more precise definition of the aspects we are considering. Third, presenting many words or statements about each box or attribute, for comparison and, perhaps, consolidation, keeping in mind explicitly the dimensionality, the units involved, the kinds of relationships, etc. Finally, formulating a series of explicit propositions about the entities and relations in these boxes, especially in columns, being ever alert for cues or hunches as to how to measure operationally, manipulate, or work with them.

Rapoport: I would like to emphasize the bifurcation of the vertical dimensions. It may go in the biological direction—species, genus, etc.—or in the social direction toward group, society, etc.

Miller: Will we need levels of biota and molecule?

Gerard: We won't often be at the level of the total biota, but it can't be left out. The emphasis is clearly going to be at the individual, dropping to organ and ascending to group and community. But the formation of autocatalytic molecules, the development of new genes and enzymes—these are the key things that operate all the way through

and the rather precise knowledge at that level should be explored or expanded. Autocatalysis is a key concept for history.

Miller: The gene is a special subsystem, which is probably larger than a molecule and smaller than a cell. The genetic problem may have to be reviewed at an intermediate level. I don't think the gene pool is a real system because it is made up of many "gene puddles" that are not contiguous in space-time, which is one of the requirements of a system as I define it. The gene pool is not contiguous in space-time because it is a conceptual term for the notion of all genes which exist within a species. Whether they covary is dependent on whether mating occurs.

Rapoport: I can't resist the temptation to point out that the gene pool is one of the most fruitful systems. To me a system is primarily a conceptualization. Contiguity in space and time may or may not be one of its aspects.

Miller: Perhaps I'll have to change my terminology. I believe it is a conceptual system, not a real system, which must be contiguous in space-time, contained within a single boundary, and with covariation among the parts. Those parts which do not covary because they are not involved in breeding are not part of such a real system.

Newcomb: Then you know only *post hoc* whether these are parts of a system or not. Have you thought about a kind of spectrum of systems, or different degrees of systemness?

Gerard: The column of hierarchies, returning to Rapoport's point, divides up each time you go through a new level. There are many organizations of molecules which don't lead to the cell, just as there are organizations of cells which don't lead to organs. And there are organizations of organisms that get you clearly into an epiorganism, like a beehive, rather than a species. At each of these, it does bifurcate, and the extent to which we are prejudicing the whole issue by selecting a particular arrangement, I think, needs more examination.

Rapoport: Another thing, a concept must be made very precise if we are going to relate it up and down. For example, you say there is stress in metal and there is stress in human beings. If you stop there, nothing is accomplished.

Gerard: Then we analyze what we mean by stress. That's really a beautiful example. Stress is an engineering term. It is used by both psychologists and physiologists incorrectly. Stress to an engineer means load. In this sense a person who is under stress means a person under a load. He reacts by developing a stress in the system.

Marquis: I think what we are all saying is that there isn't any single path or sequence of consecutive steps we should obviously follow in our discussions, that we are always going to be considering this and that, talking about definition and measurement, and at the same time being concerned about generalities.

SEPTEMBER 27, 1955

Gerard: I don't think it is really important where we start. We will probably find, when we start examining the rubrics in Table 1, that they are intimately interwoven, and the problem will be where to divide them and whether such distinctions are useful. The history of past action determines the existing architecture, which in turn determines what the possible present action and future history are. As has been mentioned by Rapoport and Marquis, you can think of language or architecture as a product of human behavior and therefore independent of it. In exactly the same way all living entities are products of past action by the evolutionary process, so you find in biology the study of organisms as they exist and the study of organisms as they came to be produced.

Miller: Our table will probably have several dimensions. For instance, how are we going to provide for the variations of individual differences? Paradigms having more than three dimensions are hard to conceptualize.

Gerard: Yes, maybe it has to be multidimensional. Well, we have before us two general lines of procedure. We can follow

either one: continue more or less at the level of dividing up the particular axis of history into more sharply defined or conceptualized units, or start with what is logically prior to that and examine the whole question of axes. If we follow our present line we have the problem of deciding which coordinates to use and which not to use, and of examining the total dimensionality or axis picture of our universe. We have discussed the level or hierarchical axis, the elements in it, and briefly their relations, including the problem of dichotomization. The other axis has to do with the properties of the systems.

Marquis: Is it possible then, considering individual differences, to think about species as the taxonomy of individuals, that is, there are some general principles about individuals, and subprinciples for types of individuals, i.e., a species? This, of course, doesn't treat a species as a system, but as a subclass of individuals.

Gerard: Why do you speak of species as a subclass of individuals?

Marquis: Well, it contains a smaller number than the total number of individuals.

Miller: That is something we should discuss, Gerard, because to me a species is a system (or org, in your terminology) only if it actually covaries. For example, those species of animals and plants which were divided, because some individuals were on top of temples in the Grand Canyon and some on the canyon floor, developed differentiation by isolation. They definitely were not covarying. They had a boundary which prevented the gene pool from intermingling, and consequently they were not a system or org. If, however, some of them had infiltrated across the boundary, they would have become part of the other system. The question is whether we are going to consider the total species, even though divided by an impermeable boundary, as a group system.

Gerard: That is exactly the problem of boundaries changing with time, as when a dividing cell becomes two, or when one community integrates into another community. These boundary problems are serious but the notion of a species is a perfectly clear and good one; as long as organisms are able to mate and occasionally do, they remain one species. If they have been kept in isolation sufficiently long so that they are no longer able to interact, they are then two species.

Marquis: I think perhaps I was suggesting that species would enter two places in the left-hand column: one as a type of individual, that is, a Manx cat and a tabby cat; the other as a type of group, in the sense of interacting and having a boundary.

Gerard: I would, I think, want to resist bringing it in under individual, if one is considering the individual as an entity. You do not at that stage examine him as a kind of an entity in a larger category. When you start doing that, then you are moving up to the next level—asking what kind of a member of the group is he? I can speak of brain cells, liver cells, etc., as different species of cells, but this is simply dividing them up for certain classificatory uses at different times. I think you are saying that at different times we may want to consider the different kinds of units in the boxes.

Miller: Take the example of two IBM computers—say the 650 versus the 704. They are different species of systems. You can program them so that they react in certain ways to certain types of influence. So a 704 that has Program A in it is in one sense a different species from a 704 with Program B. In other words, the information input which partially determines how it behaves, in addition to its basic structure which also determines how it behaves, makes a change which we cannot properly call producing a new species, but which clearly alters behavior. Concepts like *role* in social psychology or like *learning* or *instinct* in individual psychology should be reviewed in such terms. There are types of behavior which vary chiefly with genetic differences and there are other types which vary with differences in information input.

Newcomb: It is interesting, Miller, that you have talked about species always in the sense of the single individual, never mentioning collective or populational properties at all. It seems to me you are simply saying that one organism has different properties from another, and thus belongs to a different species. But, from a certain

point of view, when one organism is different from another, then it could be listed under architecture or action.

Rapoport: This would raise the question whether species as an action of an individual actually comes up as a rubric.

Gerard: Well, let's start with these questions: What about institutions? What about special orgs (a general term for material systems, including as an important subgroup living organisms or animorgs), or creations of orgs like, say, language? Is the concept of information a structural element, an action element, a history, or something that doesn't fit at all?

Miller: Gerard, you have repeatedly said that at each level the system or the org is made up of units and relationships. How are these units and relationships organized? The boundary of a system, for instance, obviously is composed of units in a specific type of relationship. Any differentiation among various sorts of elements is a consideration at the next lower level of analysis. At the moment I would hold that a concept does not exist in the sense that specific units of systems do. We have to be careful not to reify or imply inappropriate concreteness for such subjective phenomena. What is a program (which is a comparable term) in the IBM machine? It is a type of trace or engram which at a lower level may be viewed as a system. It is a gradient of electrical potential in a small area of space-time, or a set of such gradients. Isn't this what you mean when you say a person has a trace in his brain or a concept in mind? Of course this is reductionism.

Marquis: The only way you can find out about a concept is by putting in inputs and observing outputs and inferring what the program in the mind must be that would create the input-output relationship. Does it make any difference whether you know about it in terms of the program you read on a tape and put in a machine, or in terms of the reactions of a machine or organism to a series of inputs.

Miller: Either one is perfectly good science. Skinner says we do not have to know what's in the black box. We put in certain stimuli, see certain behaviors, and derive functional correlations. That is different from the other scientific method that enters the black box and tries to locate what happens. Science properly includes both methods, I suppose.

Gerard: As psychologists it is not your necessary duty to become physiologists explaining what goes on in the black box. But as psychologists you have contributed a great deal to the analysis of the "mind." So, if we consider mind as an example, does mind lend itself to division into architecture, action, and history? Are there elements which constitute significant units of analysis and communication, of structure, action, and history in the mental sphere? If there are elements, they are part of the architecture, I would think; how they operate is part of action; and how they are generated is part of history. Is this a useful way of viewing mind?

Newcomb: When you ask that question, aren't you presuming some sort of system, because system by definition means elements related in some continuous way.

Gerard: All I am asking is: What significant units does the content of psychology contain?

Miller: Can you tell us the relation between the structural and functional units in biology? That may be worth knowing since there may be a similar relation between structural and functional units in psychology.

Gerard: Well, in biology at the molecular level, the structural unit is a molecule. As for action molecules, these would be the enzyme, important for what they do rather than what they are. At another level: all the soma of an organism is a structure, the genes are action parts, they matter only in terms of what they bring about. An organism which has lost all its genes might conceivably get along perfectly well for the rest of its life. But in terms of producing another organism . . . .

Newcomb: Then their contribution to the functioning of the organism is indirect rather than direct.

Gerard: I suppose there is a time dimension in there, too. The action molecules bring about something in the future. The

difficulty arises in something like actomyosin. The muscle fiber is built of these longitudinal threads, which contain two kinds of protein molecules that are clearly structural; they remain as long as the muscle fiber does. But their role is action; they are there to make the muscle contract. Now in a way these are action molecules but they are unquestionably structural.

Newcomb: Do they have to be explicitly one or the other?

Gerard: The point is that, beyond any question, it has proven useful in dealing with living systems to separate structure and function. Organic chemistry is overwhelmingly structural and physical chemistry largely functional. Is that a useful dichotomy in the subject matter that concerns psychology? Now, what sort of units are used in psychology?

Marquis: Most of our units are relations between environmental events and consequences in human behavior. Those that refer to attributes of such relations, which might be thought of as elements of individuals, are concepts, motives, habits, affect, etc.

Newcomb: Latency would be a good example, in the case of the responding organism.

Gerard: Relation of the individual to the environment would be one category. Another would be intra-individual.

Marquis: We have specific and general relations; e.g. specific and general habits, specific and general affect. There is no sharp distinction of levels. We have a whole lexicon of words which differ principally in the generality of the stimulus-consequence relations they refer to. By that I mean we do not measure the individual or denote anything about the person in those relations. Psychological units could be grouped as follows:

1. Intra-individual
   - Idea
   - Habit
   - Affect

   (synonyms and minor modification of these terms perhaps number 300 or 400).
2. Stimulus → Consequence
   - Discrimination
   - Response
   - Latency, i.e. delay between stimulus and response.

One type of systematic effort in psychology is trying to reduce all internal concepts to idea, habit, and affect, and the suprasystem to a residue. That takes all the predictive usefulness out of them, of course, because it is in terms of specifics that we get our predictions. Specific predictions are ordinarily relations of the stimulus-consequence type, or generalized inferences from a class of stimuli and a class of consequences, with different levels of generality in our concepts which we refer to in words which are synonyms of idea, habit, and affect. There is, it should be noted, a difference of opinion as to whether motive or affect is the best way to get at the third unit.

Gerard: Motive would presumably be a synonym of affect; percept, of idea; and set, of habit?

Marquis: Set is a general habit, established with a certain time relationship to something else. Otherwise, this is the old categorization of cognition, affection, and action. No one has ever got away from that trichotomy.

Miller: Would it be constructive to say that there are two languages about behavior, one which the physiologist uses, the other which the psychologist uses? The physiologist learns about behavior by getting inside the black box; the psychologist learns about it by observing input-output relations, the input being the independent variable, the output the dependent variable, and studying the correlation between them. Idea, habit, and affect are the psychological terms. The physiological equivalent of idea is something like trace or engram, and is probably in the central nervous system. Ideas are like the figures in the problem in the electronic computer. The habit is the programming. Affect is probably not in giant brains. This is because there is a system in animals which is not in computers, an autonomic-endocrine system, which mediates affect. Certain types of stimuli, instead of resulting in ideas and habits, set off this other system which produces what we call affect or emotion, which interacts in complex ways, through the thalamocortical, cortico-thalamic, and other fibers, with the idea-habit responses, and is really

a different sort of programming. We have to recognize that it contributes important variance to the output of the behavior. Is that a good way to put it?

Marquis: That's a promising way to think of it. It raises the question then whether we should coalesce idea and habit into a single concept. The reason we don't is that one intuits (this is a subjective reason) a difference between a percept and a movement, and often the percept is a part of the change which we think is in the causation of movement.

Rapoport: Could one also put it this way, that sometimes affect has a tendency to trigger different sets of habits?

Marquis: There are all sorts of relations but they must be separated before they can be put together.

Gerard: Memory, I suppose, you would call habit?

Marquis: No, memory is in all of them. Since all of these are residues or engrams or traces or some consequence of previous experience, memory has to enter into all of them.

Miller: Memory is like latency, that is, it involves a temporal lag. Memory is the name for the fact that information is stored over time in organisms that have memories, and latency is the name for the fact that there is a much briefer temporal lag between input and output.

Gerard: Next I think we want to see if we have encompassed all the essential concepts of psychology under our headings.

Marquis: I would say the above list contains all superordinate classifications or complexes of them. For instance, "trait," which is a basic word in psychology, is a consistent tendency of a person in a class of situations to react with a class of behaviors. The trouble with our psychological units is that they are not commensurate; they do not give us any possibilities of relating one to another. We can't say that with two units of affect, the habit will be three units of something.

Gerard: We don't yet between structure and function in biology.

Newcomb: Actually most psychological work goes on at a level of intra-individual systems, more limited systems like habit and attitude. In fact, except in that fuzzy kind of study known as the case history, we never get an individual as a total system, do we?

Marquis: No, but we talk about the relations of his components, about the effects of needs on percepts, or of a particular need on a particular percept. Given a constant external environment, what are the conditions under which you get a change in a percept?

Rapoport: Those are action questions rather than structural questions. Are there any significant structural questions?

Marquis: Yes. What is the effect of establishing habit A upon the residual which we call habit B? If you change a structure, what is its effect on a neighboring structure? Is that a structural question or an action question?

Rapoport: If you are able to say that certain habits fall into a pattern, that they don't come haphazardly assorted, that there are certain classes of habits, then perhaps you can raise a structural question.

Marquis: The effect of a superordinate idea upon a subordinate idea would be structure. For instance, if you ask a person to name analogies, you are activating a superordinate idea; then you give him a specific stimulus and he gives a specific response. By that method carried out in all of its multiplicity, you get something about the structure of his superordinate and subordinate ideas.

Gerard: I think it is important to discriminate between whether or not the entity or emotion or whatever, is primarily structural, and whether or not the way you get at it is structural. Even in biological structure or in physical measurement you determine the structural positions by dynamic actions.

Rapoport: Experiments have been done in which the subjects are given terms and asked to associate others in one of two directions, either superordinal or subordinal. For example, if they are given the word "dish," the superordinal association would be household goods, subordinal, cup or something like that. It was found (I do not know how significant those findings are) that there were more superordinal associations for

mental patients and subordinal associations for the controls, with the exception of mathematicians who have a tendency toward superordinal association. I think this is a structural question. One traces the paths.

Miller: What you just said raises the question of whether "structure" means form in space, and "action," form in time.

Rapoport: We know there are intimate relations between time coding and space coding, and it may very well be that what may be time coding may reflect itself in a space coding. So it is not necessary to assume in all cases that there will be spatial relations. I would take "action" to be a term which involves a time dimension on a small scale, and "history" a time dimension on a large scale, together with the traces or results.

Gerard: I think it might be wise some time to press the psychiatrists and psychologists a bit as to the justification for choosing their classical ways of dividing their universe; it might be helpful.

Marquis: And it would be difficult; but it is worth trying. I think it is the basic question in the systemization of psychology.

To come back to our subject of relations. There is the problem of the relations of ideas to ideas and this is association. What would be the nature of those relations? We have the principles about relatedness of different ideas, e.g. contiguity and similarity. Sometimes we try to reduce them all to one, such as contiguity. Dealing with the relations of elements has been a part of psychology. I think the most promising consideration right now is the hierarchical relation of superordinate and subordinate.

One of the current interests in psychology is to discover some of the general perceptual laws, or let's say, the superordinate categories in which persons differ from each other and which in turn determine large numbers of subordinate percepts.

Newcomb: Would you agree, Miller, that contemporary personality theory has to deal mainly with the hierarchy of motives, needs, etc.

Gerard: Hierarchy in what sense?

Miller: Prepotency, dominance.

Marquis: That's the strength attribute—a result of the hierarchy. These are kinds of structural concepts. Ego, id, and superego is an alternative set of structure concepts.

Gerard: You wouldn't say id equals affect, ego equals habit?

Marquis: No, it is a different distinction.

Gerard: The value dimension might be worth examining, especially as indicated in the biological notion of health and disease, because this may be one of the main axes around which psychology and psychiatry differ in emphasis. Health and disease constitute a pair of terms which are very difficult to define.

Marquis: In a hierarchical arrangement, at the bottom there are many specific preferences, then there are general preferences, and at the top are what ordinarily are called values. This is the most general, the most embracing, way of describing all of the preferences a person has.

Miller: The other day Marquis brought up the notion that we must distinguish between real conflicts in viewpoint and mere conflict of terms. Much argument among schools of psychology concerning the ego, superego, id versus, say, idea, habit, affect, is not fundamental conflict. It can be resolved by exactly what we are trying to do here, analyzing the terms so that their relationships are known and clearly understood.

Gerard: The basis for this or that division of psychological units and the evidence for it, which we agree would be useful, should be examined. I think the value of our approach from outside psychiatry and psychology is that, each time, we must start with the assumption, which will often prove to be wrong, that the proper way to proceed is to examine this area of the universe in terms of the categories that have proved useful to other areas. In other words, to the extent that we can approach the psychological domain with biological (or other) knowledge, we should do so. This will give some basis for preferring one classification to another.

ABSTRACTS OF CURRENT LITERATURE

Prepared by Members of the Staff of the Mental Health Research Institute

American Foundation. **Medical research: a midcentury survey.** Boston: Little, Brown & Co. 1955

The report of a long study by the American Foundation, *Medical Research: A Mid-Century Survey*, is presented in two packed and thoughtful volumes. It demonstrates that solution of the outstanding problems of medicine depends upon basic research in biology, chemistry, physics, atomic energy. New knowledge and new technical facilities from these sources have changed the character of medical research, bringing the basic sciences to the forefront; and year by year it becomes clearer that direct attack on disease may be a longer way around than fundamental attack on the normal cell and protoplasm.

The first volume, *American Medical Research: In Principle and Practice*, illustrates and compares the ways in which the main purpose of each type of agency—universities and medical schools, foundations, research institutes, industries, government departments and divisions—furthers or hinders pursuit of basic research, the methods of financing research in each, and the distinctive research contribution of each type. It emphasizes the need for education in the importance of the basic sciences for progress in medical solutions.

To make clear some of the connections between basic research, however remote and however little developed, and certain medical problems is the objective of the second volume, *Unsolved Clinical Problems: In Biological Perspective*. Chapters on cancer, infertility, viruses, tuberculosis, rheumatic conditions, arteriosclerosis, hypertension, and alcoholism offer lucid and rich material in splendid review; and the final one on schizophrenia offers, as well, an integrated presentation of the interdisciplinary problems facing all concerned with behavioral science. (*RWG*).

Brodie, B. B., Pletscher, A., and Shore, P. O. **Evidence that serotonin has a role in brain function.** *Science*, 1955, 122, 968.

The data presented in this communication support the view that serotonin may be considered as a neurohumoral agent. By the use of a specific fluorometric assay for serotonin in brain, evidence has been gathered that reserpine action is mediated through the liberation of serotonin. The sedative effect of reserpine seems to be related to the change in brain serotonin rather than to the concentration of reserpine.

The hypothesis is advanced that serotonin in brain is normally present in a bound form and that, after reserpine administration, the capacity of brain cells to retain serotonin is, in part, lost. This abnormal behavior of brain cells persists even after reserpine has disappeared from the brain. Serotonin is then present in a free or physiologically active form. This free form of serotonin may be the mediator of the prolonged reserpine action. (*RT*)

Jacobs, Herbert H. **Mathematical models applied to accident processes.** In *Mathematical models of human behavior. Proceedings of a symposium.* Stamford, Connecticut: Dunlap and Associates, Inc., 1955.

The distribution of accidents in a population exposed to some environmental risk is of immediate practical interest in that it bears on the possibility of minimizing the risk by proper selection of individuals. The difficulties associated with the study of these distributions are, however, large. An experimental approach to the problem is practically impossible for obvious reasons related to investigation involving human beings. Field studies are beset by the difficulty that accidents are essentially rare events so that data samples are usually

meagre, hence the necessity of developing theoretical models which squeeze as much information as possible from data available.

The author reviews the models proposed thus far and finds them all characterized by the same shortcoming. Three factors are usually postulated to account for the distribution of accidents in a given population over a specific stretch of time: (1) Contagion effect (for which accident liabilities change with each accident experience); (2) time effect (involving changes of liability with clock time); (3) population distribution of liabilities (where the accident liability is unevenly distributed through the population). Obviously the most effective measures in accident prevention could be carried out if it were found that the population distribution had a small dispersion (indicating the presence of "accident prone" individuals). The common shortcoming of the mathematical models thus far proposed is in the circumstance that various combinations of magnitudes of the three effects give equally good fits to observed accident distributions and thus the data do not permit us to draw conclusions about the relative importance of these effects.

The author proposes a model which may enable the statistician to circumvent this difficulty. He considers accident liability at a given time to be a random variable (for which the Pearson type III distribution is assumed) and further that this variable is modified linearly by a term proportional to itself and the number of accidents which has occurred in the immediately preceding interval. From this assumption he derives the expected value of this variable as a function of the conditional expected value (given that no accidents have occurred in the immediately preceding interval) and the number of accidents which have actually occurred. Of these the first and the last are observable directly from the data, but not the conditional expected value. Attention is therefore focused on the subsamples of the population which have had no accident during some fixed interval of time. Eventually, then, a relation is derived among quantities all of which are observable, namely, the two mentioned and the expected conditional accident liability at a given time, given no accidents during the preceding interval for that subpopulation which has had zero accidents during the interval.

This relation depends on two unknown parameters, one related to the dispersion of accident liability in the population, the other to the contagion effect. The separation of the factors, therefore, has not actually been effected. The author expresses hope, however, that the model holds some promise if it becomes possible to select for study subpopulations for which the liability dispersion is suspected to be rather small. That is to say, given some a priori information about dispersion of liability, the model could confirm or dispel this hypothesis. (*AR*)

Locke, W. N. and Booth, A. D. (Eds.) **Machine translation of languages.** Massachusetts Institute of Technology, Cambridge, Mass. New York: John Wiley & Sons, 1955.

In this collection of fourteen essays, the problem of machine-made translation of languages is approached from many angles, engineering and linguistic. Designs for research and discussions of the feasibility of mechanical translation are presented, as well as descriptions of some simple methods actually in use. The relations between dictionary and machine concepts of word, the possibility of spoken rather than written, input, the problem of eliminating pre-editing or postediting, and the search for ways of dealing with translation of idioms are some of the topics discussed. (*RT*)

McClelland, C. A. **Application of general system theory in international relations.** *Main Currents in Modern Thought*, 1955, 12, 27–34.

The application of general system theory to the study of international relations is suggested as a way to achieve organic unity in the field and to reach a more satisfactory theoretical formulation. One of the basic principles of the organismic conception of general system theory, the concept of the

"open system", is offered as an alternative to the mechanistic conception predominant in the literature of international relations. It is suggested also that a more comprehensive view can be gained if the concept of equilibrium, up to now used in politics and economics, is replaced by the concept of "steady state." And finally, Bertalanffy's concept of "equifinality," stressing the importance of conditions of outflow and inflow on final outcomes, is proposed as an important organizing idea.

Under the light of the open system concept various approaches can be reconciled: the student in international relations is advised to search through the various existing concepts for significant "isomorphies," or similarities. Some examples are given of such isomorphies, by comparing current approaches, such as the "power," "communication," "decision making," and "capabilities-value-areas-field" approaches. Another contribution of general system theory to study of international relations consists in changing the focus of enquiry from the accumulation of power to the concept of organization, conceived as an "adaptive action whose purpose is the provision of specific control systems for the maintenance of a steady state in a word system whose components are nations." Finally, general system theory does not deny the dualism, ultimately emerging in the field, between action and values (or power and morals), but accepts them as useful conceptualizations gained from two different perspectives, under which two sets of isomorphically related concepts may be grouped. (*RT*)

Malcolm, D. G. **Queueing theory in organization design.** *J. industr. Engng*, 1955, 6, 19–26.

The criteria involved in setting up a group of workers to service a group of machines are discussed. Three points are covered:

1. A queueing model (not derived) regarding the length of a waiting line for service when r repairmen are repairing m machines, with a servicing factor, λ/μ.

2. Economic decision criteria detailing, under some general conditions, the cost of alternate procedures.

3. Application to organization design. A case history of such a problem.

For a more theoretical treatment of queueing theory, reference is made to *Feller, William*, **Probability theory and its applications,** Vol. 1, New York, John Wiley & Sons, 1950. (*CF*)

Marschak, Jacob. **Norms and habits of decision making under certainty.** In *Mathematical models of human behavior. Proceedings of a symposium.* Stamford, Connecticut: Dunlap and Associates, Inc., 1955.

The difference between normative and descriptive approaches to behavior are pointed out, as well as the problem of inculcating decision habits according to norms in principle accepted. The discussion is confined to situations with certain prospects, i.e., where the only decision concerns choice among alternative courses of action, the consequence of each choice being fully known in advance.

Consistency of decision is defined as satisfying the norms of constancy in time of preference among any two situations and of irrelevancy for pair-wise choices of the presence of other alternatives.

It is recognized that actual decision habits generally violate both these norms. The question is raised whether through the introduction of probability measures for choice orderings, the consistency norm cannot be re-established even in decision habits which are not consistent in the narrower sense. For example, instead of demanding that the preference ordering be always transitive—that is, $X > Y$ and $Y > Z$ always should imply $X > Z$—one can impose a "weaker" consistency: $P_{xy} > \frac{1}{2}$ and $P_{yz} > \frac{1}{2}$ should always imply $P_{xz} > \frac{1}{2}$; that is, if X is preferred to Y more often than not and Y to Z more often than not, then X should be preferred to Z more often than not.

A simple counter example leads the author to raise the question of whether there exist probability distributions of total rank orderings which satisfy the above criterion.

That such distributions indeed exist is shown by an actual example, namely a probability distribution of n real random variables $U_1 \cdots U_n$ with the following property: if a ranking r of the n possible outcomes places the outcomes in the order

X_1^r, $X_2^r \cdots X_n^r$, then the probability of the ranking equals the probability that the inequalities $U_1^r > U_2^r > \cdots U_n^r$ (where the U's are correspondingly arranged) are all satisfied.

It is then proved that this assumption not only implies consistency in the above probabilistic sense but also induces a measurement scale in the suitably defined "utilities" of the outcomes if, say F is taken to be normal with equal variances and covariances. Such a distribution has $n + 1$ parameters, or, after arbitrarily choosing an origin and scale, $n - 1$ unknown parameters, which can be determined by the $n(n-1)/2$ estimated preference probabilities (or else the hypothesis is disproved). (*AR*)

von Foerster, H. (*Ed.*) **Cybernetics.** [Transactions of the 1953 (tenth) conference]. New York: The Josiah Macy Jr. Foundation, 1955.

This is the final volume of this earliest and most unorthodox of the Macy Conferences. In contradistinction to preceding ones, no give and take discussion is included; the volume consists, essentially, of formal papers and, as an appendix, a summary of the course of discussions of the group over the span of its existence. McCulloch, chairman of the Conference, prepared this summary, as well as contributing a pleasant introduction.

The group sought "notions which pervade all purposive behavior and all understanding of the world . . . the mechanistic basis of teleology and the flow of information through machines and men." Sessions considered inverse feedback and regulation as well as goal-directed activities, from engines to societies, and emphasized information rather than energy flow, finally dealing explicitly with language. Regulatory mechanisms were examined in machines, servos; in organisms, reflex and central neuron loops; in heart; in groups of greater or lesser organization, the market and armament races. Values and computers were examined and brain mechanisms explored, with empirical as well as theoretical tools, especially looking to the problems of memory. Finally, explicit attention to information theory, to symbols, language, and communication, brings the Conference to the present session.

Grey Walter, in "Studies on Activity of the Brain," presents much of his thinking and experiment on brain function. His problem is to improve the use and interpretation of the electric indicators of brain activity, and he reports findings, obtained with his special display system (a rotating radial line seen on the tube face), indicating widespread and regular brain activity associated with visual perception and enduring as long as interest or attention lasts. He builds a theory of memory in terms of the proper interaction of two types of input streams of coded information, one closely related to sensory stimulation, the other stretched in time. Only under certain relationships are enduring traces formed.

In "Semantic Information and its Measures," Bar-Hillel summarizes work by Carnap and himself in developing a theory of information in the semantic sense, in contradistinction to information in communication, and information in the pragmatic sense. Concepts, terms, and operations are developed for a simple specimen language. Some interesting strictures emerge regarding the nature of entropy and any relation between this and the amount of semantic information.

"Meaning in Language and How it is Acquired" is discussed by Chao and, an unusual situation, is related by the author to the previous published discussions of the Cybernetics Conference. This, plus personal references and a spry style, produce a chapter of good reading, including a tour de force of a brief story told with one Chinese syllable in various uses. An analysis of the difficulty of communication, the inverse relation between rigor and content of statements, and a statement on the self-regulatory aspects of language which tend to reduce morphemes to fairly uniform length, are points of especial interest.

The volume is a worthy capstone on a most interesting conference series; one may well regret that it is a terminal one. (*RWG*)

BEHAVIOR THEORY AND THE SOCIAL SCIENCES[1]

by Charles E. Osgood

University of Illinois

The evolution of behavioral theory can be largely described in terms of what is assumed to happen between "stimulus" and "response." In any case these supposed events are generally not directly observable—they take place within a black box. Nineteenth-century psychologists filled the black box with "faculties," "motives," and "instincts." Early behaviorists refused to speculate about its contents altogether. Later behaviorists gradually introduced hypothetical "intervening variables," whose supposed nature and assumed relations became the identifying marks of various psychological theories and schools.

The recent focusing of interest on the symbolic process and attempts at constructing an "exact" theory of such processes provide an opportunity of combining in psychology the rigorous methods of natural science with the characteristic non-physical content of psychological phenomena.

PSYCHOLOGY has a rather unique position among the sciences. It is at once the most social of the biological sciences and the most biological of the social sciences. As the science of behavior—particularly human behavior—it claims and maintains active interest in much of what is equally the domain of neurology and physiology and in much of what is equally the domain of sociology and anthropology. And its practitioners reflect this Janus-faced orientation; some of us are quite indistinguishable from neurophysiologists, and indeed prefer their meetings to our own, and others of us consort almost exclusively with cultural anthropologists. Despite fractionation of both the field and the individual, this duality of interest may have certain advantages; whereas principles of human behavior may aid our understanding of problems in the social sciences, so may the nature of these problems impel reanalysis and modification of behavior theory.

[1] Originally prepared for the Conference on Analytic Systems conducted by the Department of Political Science, Northwestern University, 1955.

CERTAIN RELATIONS BETWEEN THE BEHAVIORAL SCIENCES

One indication of the validity of a principle is the vigor and persistence with which it is opposed. This holds for the rather old-fashioned notion of the hierarchy of sciences, that the principles of chemistry are reducible to physics, that the principles of biology are reducible to physio-chemistry, and so on. There is today very strong resistance to this sort of reductionism, particularly among social scientists, and they may be right, but reductionism is nevertheless a fruitful point of view. Fruitful, that is, as a guide in theorizing, not as a restriction on such activity.

Let us look for a moment at theory in psychology from this viewpoint. The psychologist, when he is functioning as a psychologist, is limited to observing what goes into the organism (which he calls *stimuli*) and what comes out (which he calls *responses*). Between these two observation points lies a no-man's-land of speculation; this region is the nervous system, also referred to these days as "the black box." Now the psychologist's task is to explain

and predict relations between and among his two sets of observables, stimuli and responses, and to do this he must make certain assumptions about what goes on in the black box. In other words, psychological *theory*, as contrasted with psychological observation, is made up of hunches about how the nervous system operates. This is true for every psychological principle, whether it be a simple association postulate by the behaviorist or a field interaction postulate by the gestaltist. These assumptions about the nervous system may be explicit or may remain implicit, but they are being made nevertheless. One criterion of psychological theory, then, is compatibility with what is known about neurophysiology; other things equal, a theory that is consistent with such knowledge is better than one that is not—but this is only one criterion.

It seems that the same logic holds for the relation between psychology and the more social sciences. In this case, behavior principles become the great unknown, about which the economist, sociologist, or political scientist necessarily makes assumptions when he postulates general principles or laws operating at his own level, explaining and predicting relations among *his* observables. When the economist states laws, for example, he is setting up some kind of "economic man," who perhaps behaves rationally and is motivated by self-interest. When the linguist describes certain general laws governing language change over space and time, he is making assumptions about how individual language users learn and modify the myriad habits of hearing and speaking. When the sociologist generalizes about status and role in a society, surely he is also assuming something about how individual human beings perceive, learn, are motivated, and come in conflict.

For the most part, as in psychological theorizing, I imagine that these assumptions about the principles of behavior, the more molecular science in this case, remain implicit. In fact, the social scientist is sometimes quite shocked at the inappropriateness of his assumptions when they are made explicit—as when labor-management people discovered that economic self-interest was by no means the only, or even strongest, motivation of workers. It would seem useful to make the assumptions of social science theory as explicit as possible so that their compatibility with what *is* known about human behavior can be examined. The greater this compatibility—again, we must stress "other things equal"—the better the theory. And let me hasten to add that this does not mean that theories at a molar level must be limited to what is known at the more molecular level. Quite to the contrary, just as the psychologist is always far beyond the neurophysiologist in the assumptions he must make about the nervous system, so is the social scientist far beyond the psychologist in the assumptions *he* must make about human behavior. Only when molar assumptions prove to be flatly contrary to what is known molecularly does the molar principle lose tenability. And, as a matter of fact, what seem to be necessary assumptions for psychological science often provide an impetus and direction to the research of neurophysiologists, and this appears to be true for the impact of the other social sciences upon psychology as well.

BEHAVIOR THEORY—AN OVERVIEW

Behavior theory, then, in the broadest sense is made up of hunches about how the nervous system operates to generate the lawful relations we observe among and between stimuli and responses. Over the past century we can trace a gradual refinement in the rules of procedure whereby psychologists make and test these hunches, and this trend toward increased rigor in theorizing has been paralleled by similar development in other social sciences.

In nineteenth-century psychology the characteristic procedure in theorizing was simply to postulate a new entity or mechanism whenever some new regularity was discovered. Whenever something needed explaining, a new explanatory device was stuck inside the black box, and it rapidly became chock-full of ill-assorted and ill-digested demons. For every nameable phenomenon of human behavior a different "faculty" would be posited to explain it; for every nameable motive, a different "instinct" would be listed as its explanation. And, at least for communicating with his patients,

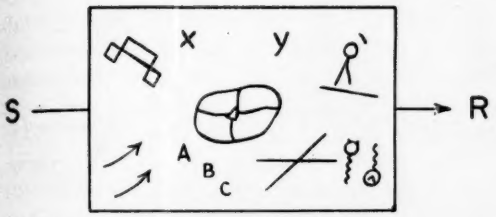


FIG. 1. The Black Box.

Freud had big, flat-footed superegos stomping around on red-slippery ids, while cleverly anxious little egos tried to arbitrate. Thus, as suggested in Figure 1, the little black box was filled with a wondrously diverse collection of explanatory devices, just about as many as there were things to be explained. This could fairly be called "junk-shop psychology." While it had the advantage of free exercise of often brilliant intuition, it had the disadvantage of complete lack of parsimony and consequent confusion.

In direct revulsion against this brand of theorizing, a group of American behaviorists around the turn of the century went to the other extreme, claimimg that the psychologist was better off if he made no assumptions whatsoever about what went on in the black box. The group included Weiss, Kantor, Watson, and somewhat later, Skinner (13). This viewpoint toward theory has come to be known as "empty organism psychology." According to this view, as shown in diagrammatic fashion, in Figure 2 there is absolutely nothing in the region between S and R, and what *is* there is none of the psychologist's business! This objective viewpoint was a healthy antidote for the loose mentalism which had preceded it, and it came to characterize American psychology. It led to increased emphasis on the details of accurately measuring stimuli and recording responses, to the establishment of dependable empirical laws relating input and output events, and to a general suspicion of unobservable explanatory devices, like "ideas," "purposes," and "feelings." But it also led psychologists to limit their interest to problems that could be handled in this simple mold—which automatically eliminated most phenomena of peculiarly human significance. The rat replaced the human as the standard subject in psychological research.

Most contemporary behaviorists could be characterized as "frustrated empty-boxers." Armed with a minimum but effective set of principles—really a set of empirical generalizations gleaned from systematic observations of S-R functions—they set out to explain and predict behavior in general. It soon became obvious, even with rats as subjects, that something had to be put back into the black box, that some intervening variables had to be postulated between S and R in order to explain our observations. But the contemporary behaviorist was more sophisticated about theory construction than his nineteenth-century forebear and, furthermore, he was under constant critical pressure from his objectively oriented colleagues. Therefore (*a*) he tried to put as little back into the box as possible, e.g., to postulate as few intervening variables as possible, and (*b*) he tried to anchor these hypothesized constructs to antecedent and subsequent observables as firmly as he could. What he has done, in effect, is to postulate an intervening response-like process which produces self-stimulation, e.g., *a two-stage mediation process*, as shown in Figure 3. In other words, "mediation psychology" sets up within the organism a replica of the S-R model, and it assumes that the same laws governing single-stage S-R processes apply to both stages of the mediation model. This is the essence of Hull's (5) "pure-stimulus-act," a response whose sole function is to provide the organism with distinctive self-stimulation (and which Hull identified, interestingly enough, with symbolic processes); it is the ubiquitous "movement-produced-stimulus" of the mechanist, Guthrie, with which he extricates himself from most tight empirical corners; and Tolman's "expectancy" or "hypothesis" in rats can, I think, be shown to be functionally the same process. Later I shall try to show how the postulation

FIG. 2. Empty Organism Psychology.

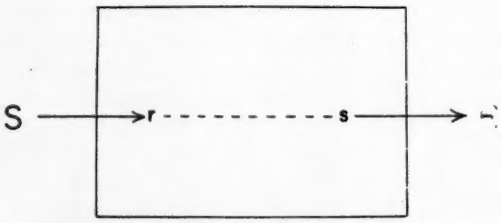


FIG. 3. Two-Stage Mediation Process.

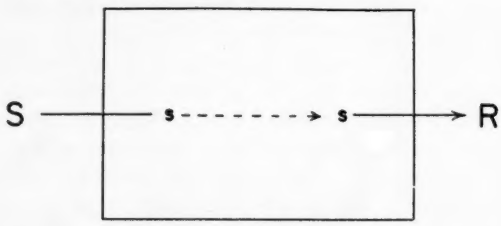


FIG. 4. S-S Psychology.

of a mediation process increases the explanatory power of learning theory.

Although the gestalt psychologists have been more concerned with perception than with learning, at the level of higher mental processes like problem solving, insight, and concept formation they have certainly come into competition with the behavioristically inclined people. Using principles of organization developed in connection with perceptual studies, they have tried to interpret processes like insight as sudden restructurings of the perceptual field of the organism. This kind of behavior theory, since it seems to involve interactions among sensory processes, has been called "S-S psychology" and might be diagrammed as shown in Figure 4. Some writers have put Tolman's (15) kind of theory in this category, because he deals with associations between signs (S_1) and their significates (S_2), but I think his view is essentially a variant of S-R behaviorism. In passing it should be noted that behavioristic S-R conceptions have had practically nothing to say about perception itself, and this is certainly an important part of behavior-in-general.

Under the pressure of trying to handle language phenomena, I have found it necessary to put something else back into the black box. Along with symbolic, representational processes (which I think can be handled with the two-stage mediation model), language behavior also includes

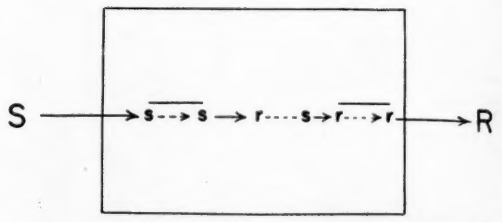


FIG. 5. Central Integration Among Neural Events.

the most complex integrations of both perceptual processes and motor skill processes. On both input (decoding) and output (encoding) sides of the equation, as shown in Figure 5, I think we must postulate *central integration* among neural events. We need something of this kind to handle standard perceptual phenomena like closure as well as standard motor phenomena like the execution of skilled movements at a rate far greater than could be explained by S-R feedback mechanisms. We might call this model "integration-mediation psychology." But we have been leaping ahead much too quickly here; let us back-track a bit to the simple "empty organism" model and see what basic principles appear to govern S-R relations.

BASIC PRINCIPLES OF S-R BEHAVIOR THEORY

Although learning theorists disagree among themselves somewhat as to the language to be used, and occasionally as to the necessity of certain postulations, the basic principles described below would find pretty wide acceptance. They all deal in one way or another with the formation of *associations* among input and output events by organisms and thus reflect our heritage from philosophical Associationism—except that the events associated are stimuli and responses rather than ideas. They all may also be viewed as fundamental assumptions psychologists make about how the nervous system both organizes and is organized by behavioral experience.

1. Contiguity. In general, *the closer in time the occurrence of a stimulus event and a response event, the greater the increment in their association*. It is the time interval between a stimulus and a response, not two stimuli, with which this principle is concerned. Although it is true that in *condi-*

tioning procedure we vary the time interval between the to-be-conditioned (or to-be-associated) stimulus (CS) and the unconditioned (or already-associated) stimulus (US), this is simply to guarantee sufficient contiguity between CS and the R which US controls. Some such principle as this is necessary on common-sense grounds; all events that occur in a lifetime are not associated indiscriminately. Its operation is also clearly evident in all experimental data on learning; the more we separate in time the S-R events to be associated, the more difficult it becomes to form the association, e.g., to get this S to elicit this R predictably. Degree of associativeness is probably a negatively accelerated decreasing function of time interval and limited to short time periods in the pure case. A number of assumptions about the nervous system are implicit in this psychological principle: (*a*) that the neural correlates of motor events are capable of becoming dependent on those of sensory events; (*b*) that these neural correlates must persist or reverberate over at least short intervals of time; and (*c*) that to become "associated," the sensory and motor correlates must be simultaneously active and somehow brought into propinquity in space. The latter notion really depends upon the underlying assumptions of materialistic, natural science.

2. Summation. Common-sense observation reveals certain consistencies and predictabilities in the stream of behavior: when the postman approaches the door, your dog can be expected to bark; when someone gives your child a present, he can be expected to say "Thank you." We call such predictable regularities *habits*. We also note that these habits can vary in *strength*, which we index by observing the probability with which R follows S, the amplitude of R given S, and the latency or speed with which R follows S. Furthermore, this *habit-strength* seems to increase with repetition or practice. As a general principle we may state that *successive increments in the association of a stimulus event with a response event summate to yield habit strength*. It seems reasonable that the strength of any habit should have a limit or asymptote, and empirical data do show that with large numbers of repetitions the increments per trial become imperceptible. Habit strength is an increasing negatively accelerated function of the number of repetitions. Here again the psychologist makes assumptions about neurophysiology—perhaps that at the synapses linking the neural correlates of sensory and motor events certain progressive and cumulative changes in conductivity can occur.

3. Generalization. Now, it is one of the fortunate things about behavior as we observe it that having learned to make a particular response in a particular situation we do not have to completely relearn this habit if the situation changes slightly, nor do we have to accomplish the behavior with precisely the original movement. Having learned to say "Thank you" in the home, your child is likely to employ the same social grace elsewhere; having learned to walk normally, we can usually hobble about pretty well with a sore ankle. It appears that habits spread or generalize among similar stimuli and similar responses. There is also a logical necessity for postulating a generalization principle.

Our habit principle says that successive increments in the association of a stimulus and a response summate; yet if there is one thing we know about behavior it is that precisely the same stimulus and response are never repeated. How, then, can there be summation among unique events? The following principle takes account of common-sense observation and the apparent paradox: *the habit-strength generated between a stimulus event and a response event generalizes to other stimulus and other response events, the amount of such generalized habit strength being a function of* (a) *the similarity between directly associated events and non-associated events*

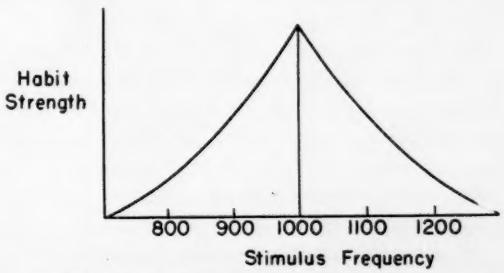


FIG. 6. An Example of Generalization.

and (b) *the strength of the original association.* Laboratory evidence is consistent with such a principle; if we train a rat to jump in the air to a 1000-cycle tone, by pairing this tone repeatedly with shock, and then test with other tones of varying frequency, results like those shown in Figure 6 are typically obtained—the more similar the test tone to the conditioning tone, the greater the amplitude (speed, probability) of generalized jumping. This principle assumes that, in both sensory and motor nervous systems, the neural correlates of similar events are closer together than those of less similar events, and both sensory and motor projection systems do seem to display just such an organization.

4. Motivation. The mere fact that a set of associations or habits have been learned does not guarantee that they will be utilized. Place a satiated rat at the starting point of a well-learned maze and he will probably groom himself or go to sleep rather than running; place a highly motivated worker on the job and he will usually produce more than a poorly motivated worker. Learning theorists phrase this general law of behavior somewhat as follows: *motivation combines multiplicatively with habit-strength to yield performance.* In other words, if either the motivation level or the learned habit-strength is zero, the probability of the behavior in question becomes zero; and further, this implies that within limits, at least, we can trade or exchange motivation for learning and maintain a constant output —a highly motivated student with poor training may equal in performance a poorly motivated student with good training. In the laboratory we usually index drive level operationally by the number of hours' deprivation of food or water in the case of hunger and thirst drives, by physical intensity of shock in pain drive, by the measured amount of testosterone in the blood in sex drive, and so on; with human subjects we usually depend upon such indirect indices as the subject's verbal statements, what he tells us about himself, or even, quite circularly, upon how energetically he performs.

5. Reinforcement. It is commonplace in the home as well as in the animal laboratory that subjects learn better when they are rewarded than when they are not. In fact, in the laboratory as well as outside of it, unrewarded or unreinforced associations tend to get weaker despite repetition, a process we call *extinction.* Some psychologists, so-called "effect" theorists like Thorndike (14) and Hull (6), have incorporated this basic empirical fact about behavior in a postulate; others accept the phenomenon but think it can be derived from other principles, for example, a sheer contiguity theorist like Guthrie (3). Since this paper is not the place for arguing fine points of theory—and, indeed, the amount of evidence that would have to be evaluated is prohibitive—I shall merely state the reinforcement principle in its usual form: *the size of the increment in association between a stimulus event and a response event varies inversely with the time interval between association and a reinforcing state of affairs and directly with the amount of reinforcement.* If a rat is learning to press a bar in order to obtain food pellets, (*a*) the greater the delay in receiving pellets the more trials it takes to reach a criterion of performance and (*b*) the smaller the size of the pellet the longer it will take. What actually constitutes this "reinforcing state of affairs"? Some theorists, like Hull, specify reduction in the intensity of a drive (hunger, thirst, pain, anxiety, and so on) as the essential condition. A possible neurophysiological correlate of reinforcement has appeared in some recent experiments at McGill—when performance was accompanied by electrical stimulation of a certain brain region in animal subjects, learning occurred; when stimulation of this region was eliminated, learning failed to take place and old habits were extinguished. What relation this may bear physiologically to drive states and their reduction I do not know.

6. Inhibition. Now, let us look at two paradoxes. The first, already mentioned briefly, is *extinction.* Suppose a rat has already learned to press a bar to obtain food pellets that roll down into a cup in what we call a Skinner box. Suddenly, without the rat's knowledge, we shut off the supply of pellets. For some period of time, the rat will persist in pressing the bar

because he has both a learned habit and is hungry; but the rate of pressing will gradually go down until he fails to press the bar for some criterion interval, say 5 minutes, and we say he has extinguished. This in itself is paradoxical, because there is nothing in our principles so far to account for this behavior—a motivated organism is reacting to a stimulus (pressing the bar) repeatedly, S is being followed by R, yet the habit seems to be getting weaker. Why, you may ask, do we not simply postulate that unreinforced associations are weakened by repetition? If we do this, we run into another paradox, called *spontaneous recovery*. Let us now take our extinguished rat out of the Skinner box and let him rest overnight in his home cage; the next morning we put him back in the Skinner box again and, lo and behold he is busily pressing the bar again. If we claimed that the habit had been lost during the previous extinction, then it would appear that a habit could be learned spontaneously without reinforced repetitions, which would be contrary to our principles! These paradoxical phenomena are well documented, and it is situations like these that "put the pressure" on theorists.

To handle this type of situation we may postulate an inhibition principle: *the execution of any response produces an increment of inhibition toward making that response, such inhibition increasing with the effortfulness of the response and dissipating spontaneously with rest.* During learning the effects of reinforcement counteract these work-produced decrements, unless the rate of responding is too great, but as soon as reinforcements cease to be given, inhibitory tendencies mount to a point where the animal stops responding, i.e., extinction. During rest in his home cage, the work-produced inhibition dissipates and in the morning the learned habit appears again spontaneously. But haven't we explained too much and gotten ourselves into another paradox? If what we have said were literally true, animals would never be able to unlearn habits, and this doesn't jibe with facts—the recovery from successive extinctions actually becomes less and less. As a matter of fact, we can handle this without any new principles. Since the fatigue sensations accompanying work-produced inhibition constitute an unpleasant drive state, and since rest involves spontaneous reduction in this drive state, we have the necessary reinforcement conditions for new learning. During the later stages of extinction, whenever the subject stops reacting and rests while still in the situation, he is actually learning *not* to press the bar and this response-of-not-responding is being reinforced by fatigue-drive reduction. This inhibitory habit subtracts directly from the previous excitatory one and constitutes the permanent unlearning produced by extinction operations.

7. Selection. We are all familiar with situations involving response conflict. Approaching a divided staircase, we experience a momentary ambivalence as to turning left or right; warned not to touch the freshly baked and frosted cake, but seeing it right there before our eyes, we can remember oscillating between approach and avoidance. In general, *whenever two or more responses have been associated with the same stimulus, the reaction having the momentarily strongest habit strength will occur.* If the reactions in question are compatible, of course (like smiling and shaking hands), then both may occur simultaneously. When a number of incompatible reactions are associated with a common stimulus, we speak of a *divergent hierarchy*; this is a response competition situation, and probabilities of alternatives depend on contextual factors and the strengths of motives associated with each. When a number of stimuli are associated with a common reaction, we speak of a *convergent hierarchy*; this is a facilitative, transfer situation and follows the principle of generalization already discussed.

These seven principles are by no means exhaustive, they are not worded precisely as any particular theorist might state them, and they are not the exact assumptions made by any particular theorist, but they do seem to represent a set of basic notions which are common in one form or another to most psychologists who deal with learning. Let us analyze a few standard learning situations.

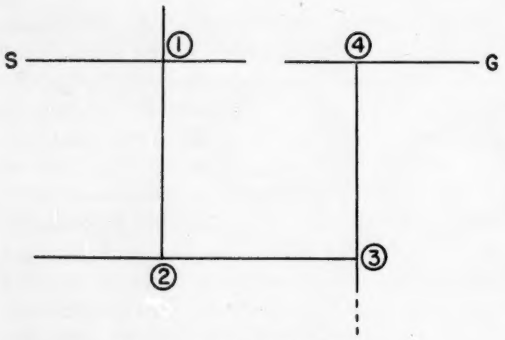


FIG. 7. A Multiple T Maze.

Maze learning. The maze is one instance of a general category we refer to as *instrumental learning*; the subject must make the correct series of responses before reinforcement is given, and hence the correct responses are "instrumental" in obtaining reward. Figure 7 diagrams a multiple-T maze, where *S* is the starting point, *G* is the goal box where a food reward is found, and the encircled numbers refer to successive choice points. At each such choice point we have a stimulus situation and two alternative reactions, turning left or right. We shall assume that our subjects are highly motivated. Now we may ask a series of questions of Nature and predict from theory what her answer will be. (*a*) *What will be the order of eliminating errors?* Prediction: other things equal, errors will be eliminated first at the choice point nearest the goal and work back progressively toward the starting point. Why? The closer an S-R association to the point of reinforcement, the greater the increment in habit-strength. Although the animal always gets to the goal, any wrong response necessarily causes a longer delay in obtaining reward, and hence the ratio always favors the correct choice. (*b*) *How will increasing the physical similarity between choice points* 3 *and* 4 *affect errors at choice point* 3? Prediction: increased similarity will increase errors at choice point 3. Why? In (*a*) above we have deduced that the right-turning response at 4 will tend to be learned earlier than the left-turning response at 3; the greater the similarity of choice points 3 and 4, the greater will be the generalization of the right-turning response from 4 back to 3, where it is an error. Since this "error" is also followed by reinforcement, it will be strengthened and compete more successfully with the correct reaction at 3. (*c*) *What effect upon errors at choice point* 3 *will lengthening the blind alley there have*? Prediction: the longer the blind, the more quickly it will be eliminated. Why? Two principles combine to yield this prediction: in the first place, the longer the blind the greater the delay in reinforcement and hence the more favorable the ratio of strengthening for the correct reaction; in the second place, running a long blind produces more extinctive inhibition, which would be used to explain the fact that the rat runs less and less far into the blind as trials continue.

Classical conditioning. Pavlov (12) studied a learning situation in which one stimulus, which was not originally capable of eliciting a response, became gradually capable of doing so if it was paired with another stimulus which regularly and reliably produced this response. Figure 8 diagrams the relations assumed in conditioning in two ways, first descriptively (*A*) and then interpretively (*B*) according to the principles we have been discussing. A typical case would be that in which a dog subject learns to salivate to a tone (CS), food powder blown into the mouth being the unconditioned stimulus (US) and amount of saliva secreted being the reaction measured (R). The model given as *A* makes it appear as if this were a different "kind" of learning, simple substitution of one stimulus for another. The model given as *B* assumes that the dog must be hungry (motivated) in order to be conditioned and that the US (food powder) has a dual function, both in reliably eliciting the correct response of salivating (an unconditioned reflex in this case) and in providing reinforcement (re-

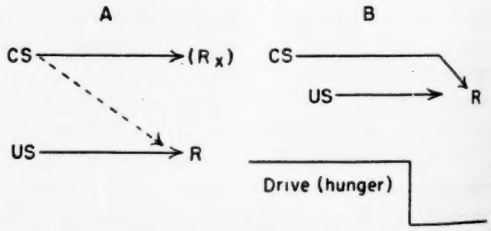


FIG. 8. A Diagram of Classical Conditioning.

duction in hunger drive). Thus in conditioning, according to this view, we have contiguity between CS and R, motivation, reinforcement, and evidence for summation into habit-strength, in that the probability of the tone eliciting salivation is shown by experiment to increase with the number of repetitions. We may ask some additional questions: (*a*) *Does ease of conditioning vary with the time interval between CS and US?* Prediction: from some optimal interval (about .5 sec., presumably dependent on nervous system functioning), ease of conditioning will decrease as interval between CS and US (and hence interval between CS and R) increases. This phenomenon is regularly observed. (*b*) *What will happen as the pitch of the tone is varied?* Prediction: the less similar in frequency any test tone from the tone used during conditioning, the less will be the amount of saliva secreted (generalization). This phenomenon is regularly observed. (*c*) *What occurs if the US is discontinued?* Prediction: since reinforcement is no longer given, but the reaction keeps occurring, inhibition develops and the magnitude of the salivary reaction will decrease. This is known as extinction of a conditioned reaction.

Discrimination. Now, suppose that in the situation we have just been describing one tone is always associated with the giving of food powder (S_A) and another tone is always associated with the withholding of food powder (S_B). As shown in Figure 9, the habit-strength associated with the positive stimulus must generalize along the pitch continuum. Similarly, we assume that the inhibitory tendency associated with the negative stimulus must generalize (e.g., generalization of a response of not-responding). If we apply principle 7 (selection among competing R's according to momentary habit-strength) to each point along the continuum, we in effect plot the resolution of these two gradients (dashed line). Wherever the resolution is above the base line, there will be that degree of excitatory tendency; wherever the resolution is below the base line, there will be that degree of inhibitory tendency. What this operation does in effect is to narrow the range of generalization, i.e., produce discrimination.

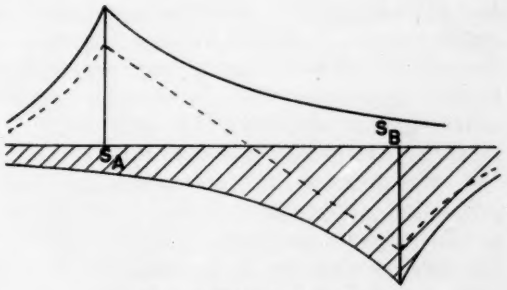


FIG. 9. Discrimination.

As can be seen, it could be predicted that the more similar the two tones, the more difficult it should be to obtain discrimination, and this is precisely what is observed. To climb out into the world of everyday affairs again, it would be predicted, for example, that a man hired to reject unpainted plaques in an assembly line would have a harder job if the paint were white than if it were red.

INSUFFICIENCY OF THE SINGLE-STAGE S-R MODEL

Employed with ingenuity, even such a minimal set of tools as these can be effective in predicting and interpreting over a considerable range of behavioral phenomena. But there are some gross insufficiencies with this single-stage S-R model, particularly when one tries to handle typically human activities. For one thing, the entire problem of *perceptual organization* seems to be omitted, since it involves relations among input events, S-S integrations; for another, the development and use of *motor skills* is omitted, to the extent that such behavioral phenomena involve R-R integrations; and finally, in its unextended form, the single-stage S-R model is incapable of handling *symbolic processes*, which are so characteristic of human behavior. The reader should keep in mind, however, that in pointing out these insufficiencies of the S-R model I am not claiming that its basic principles are thereby invalid—to the contrary, in dealing with the symbolic processes, at least, the same general principles will be assumed to operate.

Perceptual organization. Our sensory, receiving systems are so planned by Nature that points on the receptor surface (e.g., the retina of the eye) are mapped in more or

less point-for-point correspondence onto the sensory cortex in the brain. In other words, activity in certain sensory portions of the brain is isomorphic with the stimulus events affecting our receptors. This first stage in reception is called *projection*. Furthermore, the transmission characteristics of this projection system have been shown to be unaffected by experience, e.g., unmodifiable through learning. But it has also been shown that beyond this first projection level, point-for-point correspondence breaks down, and organization on the basis of learning replaces organization on the basis of innate "wired in" isomorphism. What is the nature of this "perceptual learning"? A very simple principle of *sensory integration* on the basis of sheer frequency can be postulated here (cf., Hebb, 4): *the greater the frequency with which stimulus events* A *and* B *have been paired in input experience, the greater will be the tendency for the central neural correlates of one*, a, *to activate the central correlates of the other*, b. If the frequency of stimulus pairing is very high, the occurrence of one may become a sufficient central condition for evoking the central correlate of the other (evocative relation); if the frequency of stimulus pairing is less, the occurrence of one will merely "tune up" or prepare occurrence of the other (predictive relation). This principle obviously makes certain assumptions about the nervous system, specifically, that nerve cells more central than the projection system which are in fibrous contact and are caused to be simultaneously active somehow increase their capacity to fire one another.

One behavioral manifestation of the working of this sensory integration principle can be seen in the well-documented phenomena of *closure*. We typically inspect our environment with rapid samplings and the sensory information must be sketchy, yet we perceive wholistic, complete forms; experimentally, an incomplete circle can be flashed on a screen briefly and most observers will report seeing the completed form—the given signals are sufficient to evoke the central correlates of those which are missing in a familiar form. Similar effects have been observed in language decoding: when printed words are flashed briefly on a screen, a highly regular inverse relation is found between the duration of the flash necessary for perception of a word and its frequency-of-usage in printed materials like magazines and books; the same relation is found for the auditory intelligibility of spoken words. In general, on the sensory side of behavior, it appears that the original chaos of sensations which William James postulated for the newborn infant is gradually molded into a world of perceived forms on the basis of sheer frequency of stimulus pairing in experience.

Motor skill. On the final output side of behavior we also find a "wired in" isomorphic projection system between motor cortex and muscle elements; this motor projection system is also unmodifiable through experience. But just as in the sensory system, the nervous tissues in the motor cortex more central to the motor projection system seem capable of being organized on the basis of sheer frequency of response pairing and sequencing. We can formulate a principle of *motor integration* or programming: *the greater the frequency with which response events* A *and* B *have been paired in output experience, the greater will be the tendency for the central neural correlates of one*, a, *to activate the central neural correlates of the other*, b. Again, extremely high frequency of response pairing or sequencing will result in completely evocative units (like the syllables in spoken language) and lesser frequencies will result in merely predictive integrations (like the grammatical sets or expectancies in language). S-R behaviorists have usually tried to explain motor skills like shoe tying, typing, and piano playing in terms of proprioceptive feedback—each muscle movement in a sequence produces proprioceptive self-stimulation which becomes the cue for the next response in the series—but measurements on the maximum rate of nerve fiber conduction show that there just simply isn't enough time for such control in rapidly executed skills like playing a cadenza.

The merely predictive sensory and motor integrations are interesting because they provide an experience-based stability in both decoding (receiving) and encoding (expressing) activities of the organism.

Sight of a falling object is predictive of a certain striking noise; hearing Lowell Thomas say "So long. . ." is predictive of hearing ". . .until tomorrow"; seeing a piece of ice and one's hand reaching toward it is predictive of cold sensations—and in all such cases one's shock of surprise is great when these predictions occasionally are not borne out (indeed, if the lack of confirmation is not clear and intense enough, we often misperceive in accordance with what is expected). On the motor side, lighting a cigarette is predictive of blowing the match out; saying "*either* he will. . ." is predictive of saying "*or* he. . ."; unbuttoning one's shirt is predictive of peeling it off—one is usually aware of a sense of incompletion when, for some reason, such integrated sequences are interrupted. In other words, there is a syntax of perceiving and behaving just as there is for hearing and speaking a language, and this syntax seems to depend on the formation of S-S and R-R associations or integrations, as I have called them.[2]

Symbolic processes. The unextended S-R model uses certain hypothetical constructs, like habit-strength and inhibition, to make possible summary statements about a wide range of phenomena, but it does not use intervening variables—its principles concern relations between external S and external R. There are many observations we make of stimulus-response relations, however, which require the postulation of intervening variables; these observations fall generally in the category of symbolic or representational processes and, although they are certainly more characteristic of human behavior, they can also be observed in primitive form in subhuman species. Before we try to extend our principles to cover such phenomena, let us review some of the evidence that seems to require such extension.

We may look first at some evidence in rat behavior. In the typical *avoidance training* situation a buzzer sounds a few seconds before a strong shock is delivered to the animal's food-pads; in the course of reacting agitatedly to shock, the rat hits upon the correct response, perhaps turning a ratchet-wheel which eliminates the shock; after some number of trials, the rat will come to react to the *buzzer* in a way that takes account of the coming shock, e.g., will turn the ratchet-wheel before the shock comes on. In an operational sense, we could say that the rat has foresight, that it anticipates the shock, that the buzzer has become a *sign* of shock. To demonstrate that the buzzer has indeed acquired a danger significance for the rat, we can now remove the ratchet-wheel and substitute a hanging chain; without ever being shocked at all in this new situation, the rat can learn to pull the chain if that act turns off the buzzer; he will be galvanized into vigorous activity by the persisting buzzer (*acquired or symbolic drive*) and will learn the new act on the basis of eliminating this anxiety-producing sound (*acquired or symbolic reinforcement*). Perhaps more striking is this demonstration of "intelligent transfer" in the rat: Initially these animals are subjected to unavoidable shock in a little pen, the experimental group having a buzzer paired with the shock and the control group also hearing the buzzer but not paired with the shock—all they can do is freeze and crouch in this situation; then both groups are trained to escape a shock in a different apparatus by running and leaping over a barrier, until they do this immediately when they feel the shock come on; finally, both groups are tested by suddenly sounding the buzzer—the experimental animals immediately run and leap the barrier *to the buzzer* whereas the control animals do not. Since the buzzer-sound has never been single-stage associated with running and leaping, we must infer that the buzzer excites *some intervening process* which is associated with this response, e.g., some shock-representing process. Many other sources of evidence from animal behavior could be cited, delayed reactions when cues are eliminated, problem solving, learning from sheer observation, and so on, but this will have to suffice.

Nearly all the phenomena we consider peculiarly human depend upon symbolic mechanisms. A simple but typical example

[2] A more detailed analysis of such sensory and motor integrations is given in a paper by this writer prepared for the Colorado Symposium on Cognition (10).

is *semantic generalization:* a subject is shocked whenever the word JOY appears among lists of other words; some response, such as finger-flexion or electrodermal skin reaction, is recorded; then a list of test words is presented, and we find that the subject makes a generalized reaction to words like GLEE and FUN, but not to words like BOY or TOY—and this despite the fact that the latter are physically more similar to JOY than the former. It is clear that some symbolic, meaningful process, common to JOY and GLEE but not to JOY and BOY, must be mediating the transfer of reaction. Another impressive demonstration of the necessity of postulating a two-stage process in human learning, when you think of it, is the simple fact that we humans can be tremendously modified in our attitudes and subsequent behavior by merely watching a TV program like the Army-McCarthy hearings—yet at the time S is given no observable R is made. Similarly, a man can spontaneously initiate complex instrumental responses, smiling, walking out of his study, talking about a thousand-and-one things, without any observable appropriate S being present. And needless to say, the entire structure of language behavior, upon which any adequate understanding of social science must be based, is quite incomprehensible when viewed within a single-stage S-R framework, because within such a framework questions of significance, meaning, and intention can only be sidestepped.

REPRESENTATIONAL MEDIATION PROCESSES IN BEHAVIOR

We have come upon a number of facts about both animal and human behavior which cannot be handled with a single-stage S-R model. This means that we must make some additional assumptions about what goes on in the black box. In keeping with the law of parsimony, we wish to make as few assumptions as possible, and yet be able to incorporate the relevant phenomena. In order to account for S-S relations (perceptual organization) and R-R relations (motor skill), it seemed possible to cover the ground by simply postulating neural integration among the central correlates of sensory and motor signals respectively on the basis of sheer frequency of input or output pairing. What is the least amount of additional baggage needed to handle symbolic processes? In effect, we shall break our single-stage process into two stages by postulating a learned, self-stimulating, response-like process as an intervening variable. The first stage, which we shall call *decoding*, will be the association of external stimuli with these internal, symbolic responses; the second stage, which we shall call *encoding*, will be the association of the self-stimulation (produced by these symbolic reactions) with external responses. The great advantage of this solution is that, since each stage is itself an S-R process, we can simply transfer all the conceptual machinery of single-stage S-R psychology into this new model without new postulation. The critical problem, of course, is to anchor this intervening variable firmly to both antecedent and consequent observables, e.g., objective stimulus and response variables. This mediation theory is treated more fully elsewhere. (9).

Development of representational mediators

We start with the fact that certain stimulus patterns have a "wired in" connection with certain behavior patterns (unconditioned reflexes) and additional stimuli have acquired this capacity (conditioned reflexes). Food powder in the mouth regularly and reliably elicits a complex pattern of food-taking reactions (including salivating, swallowing, and the like); a shock to the foot-pads regularly and reliably elicits a complex pattern of escape reactions (leaping, running, urinating, autonomic "fear" reactions, and the like). What I shall call a *significate*, then, is any stimulus which, in a given situation, regularly and reliably produces a predictable response pattern. For the naive organism, there are multitudes of stimuli which do *not* have such characteristics—a buzzer-sound does not reliably produce escape behavior like the shock does; the sound of a metronome does not reliably produce food-taking behavior like food-powder-in-the-mouth does. And for the human infant, the visual image of its bottle or the breast does not initially

produce sucking, salivating, and the like. Our problem really is this: how can neutral, meaningless stimuli like these become meaningful *signs* for the organisms affected by them?

A moment's thought convinces us that ordinary single-stage conditioning does not provide a satisfactory answer—the reactions made to signs of objects (CS) are almost never the same as those made to the objects themselves (US), even for the rat subject. When the buzzer sounds in the avoidance situation, the rat behaves in such a way as to take account of the shock, not as it would to the shock itself; when the human hears the word FIRE, he behaves in a way appropriate to the whole situation including this sign, not as he would to the object referred to. But if we look into the so-called conditioning situation more carefully, we can, I think, see an answer to our problem. Many experiments on the details of the conditioning process combine to support the following conclusion: not all components of the total unconditioned reaction are conditionable to, appear in, the conditioned reaction equally; rather, certain "light-weight" components, like autonomic reactions (changed heart-rate, glandular secretion, etc.) and minimal muscular and postural adjustments (tensing, lip-licking, etc.) tend to be elicited by the CS long before the "heavy-weight" components like paw-lifting, chewing, running and the like. Incidentally, this could have been predicted from single-stage principles—"heavy-weight" reactions will involve greater amounts of work-produced inhibition.

The diagram in Figure 10 represents the formation of a representational mediation process, which we may phrase as a principle in the following way: *Whenever a neutral stimulus is contiguous with a significate, and this occurs sufficiently close to a reinforcing state of affairs, the neutral stimulus will acquire an increment of association with some portion of the total behavior elicited by the significate as a representational mediation process.* This stimulus-producing response process (r_m) is *representational* because it is part of the very same behavior produced by the significate itself (R_T)—thus the buzzer becomes a sign ($|\overline{S}|$) of shock ($\mathring{S}$) rather than

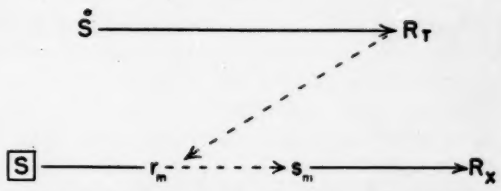


FIG. 10. Formation of a Representational Mediation Process.

a sign of any of a multitude of other things. It is *mediational* because the self-stimulation (s_m) produced by making this intervening reaction can become associated with a variety of instrumental acts (R_x) which "take account of" the object signified—the anxiety state generated by the buzzer may serve as a cue for turning the ratchet which eliminates the shock.

Now, with this principle in mind, let us look back at some of the symbolic phenomena which proved embarrassing to the single-stage conception. In avoidance training, shock reliably produces total escape behavior; being regularly paired with buzzer, or the traces of the buzzer sound, this initially neutral stimulus acquires some distinctive portion of the total behavior to shock (here, what we call an *anxiety reaction*); the self-stimulation from this mediator also becomes associated with the instrumental act of ratchet-turning. That this mediator has an impelling danger significance (symbolic punishment) is shown by the fact that a new instrumental act, like pulling the hanging chain, can be learned simply on the basis of turning off this stimulus. We could fairly say that this buzzer-sound has acquired a new *meaning* for the rat, has become a sign of something other than itself. The learning of meanings is central in human behavior, of course, so we look at semantic generalization in terms of this model. Prior to the experiment itself, we assume that our subject has learned the meanings of JOY, GLEE, BOY, and so forth. JOY and GLEE, as arbitrary social signs, occurred in similar total situations, in which intense pleasant reactions were taking place; therefore these signs will be associated with representational portions of these reactions. BOY will be associated with quite different mediation processes, we assume. Now, since the quality of self-stimulation depends

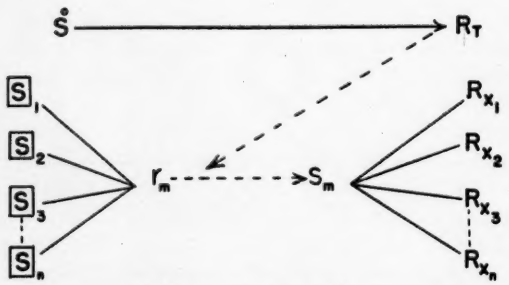


FIG. 11. Mediating Role of Representational Process.

entirely upon the reactions producing it, it follows that the s_m produced by JOY (and conditioned to a new reaction in the experiment) must be more similar to that produced by GLEE than that produced by BOY; therefore it follows from the generalization principle (applied to our stage 2, encoding) that the new reaction conditioned to JOY will transfer more to GLEE than to BOY.

As illustrated in this last analysis, the most important role of representational processes in behavior is as the common element in mediated generalization and transfer. It is this, I think, which gives the tremendous flexibility we observe in the behavior of higher animals, particularly humans. Figure 11 diagrams this mediating role. Whenever a set of stimuli are associated with the same significate, and hence with the same total behavior pattern, they must become a hierarchy of signs associated with the same representational process and hence have the same significance. Thus, for the rat, the sight and smell of the food pellet, the cup in which it is found, the click which announces its delivery, and so forth all become more or less equivalent signs of food reinforcement; thus, for the male human, a lace handkerchief, the smell of perfume, words like "pretty" and "nice" become partially equivalent signs of the feminine goal object. Similarly, on the other side of the behavioral equation, this common mediation process as a form of self-stimulation becomes associated with a hierarchy of overt instrumental acts. To situations which have in common a danger significance, the rat has learned to run, to crouch and freeze, to squeak, to turn a ratchet, and so on; in similar situations, the human has learned to run also, to plead verbally, to "talk his way out," and so forth. Selection among such a divergent hierarchy of responses depends upon both the absolute habit-strength of the alternatives and the contextual cues present (e.g., the human is more likely to "talk his way out" when in a locked room than when in an open field). Note the remarkable increase in flexibility such a mechanism provides: Once a novel stimulus has been associated with an established mediator (e.g., a flickering light is made a sign of shock), it immediately takes command over the entire hierarchy of instrumental acts already associated with this mediator (e.g., the rat will operate the ratchet to turn off this flickering light without new training). Similarly, once a new instrumental act has been associated with an established mediator, it immediately becomes available to all members of the class of signs having the same significance.

This analysis leads one to draw a distinction between what might be called primary and secondary learning. *Primary learning* would involve the original formation of sensory integrations (perceptions), motor integrations (skills), and representational processes (meanings); such learning is the daily work and play of the growing child, and what evidence we have indicates that this primary learning is extremely time-consuming and laborious (albeit happily dealt with as natural "play" for the most part). *Secondary learning* would involve nothing more than associative reshuffling of signs with mediators and of mediators with instrumental acts. This kind of learning is characteristic, I think, of the mature or

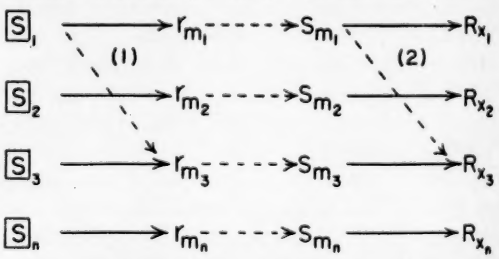


FIG. 12. Secondary Learning: (1) Change in Significance of a Sign. (2) Change in Instrumentation.

sophisticated organism. One type of secondary learning, as shown in Figure 12, is *change in significance of a sign*: Most so-called conditioning experiments are of this sort; a dog for whom a whistle has already some significance (seldom investigated, actually) must come to associate this stimulus with a food-taking significance, indexed by salivation. A young lad for whom an officer's uniform has one significance, based on movie-going experience, comes to modify this significance somewhat after being drafted (in passing, note the relation of this analysis to status and role conceptions in sociological theory). Another type of secondary learning is *change in instrumentation*. As we have seen, the rat may be induced to modify its overt responses to a buzzer-sound (from running to ratchet-wheel turning) without changing the previously learned danger significance of this sound. The maturing human changes radically his instrumental responses to his mother and father, to his peers and superiors, as he shifts through childhood, adolescence, and maturity, without necessarily changing the significance of these people to him; similarly, he learns to behave differently to danger signals, food signals, and so on; on the other hand, the changes in his behavior to females of his species probably involve simultaneous changes in both significance and instrumentation. What is suggested, in other words, is that the average mature individual (human or animal) engages in very little "new" learning, very little in the way of forming new perceptual integrations, new skills, or even new meanings—he has the familiar physical and social world pretty well organized with respect to himself —rather, what learning he does do is concerned with occasionally shuffling relations among established perceptions, meanings, and skills, and most of us do precious little of this.

This leads rather directly into the problem of *psychological units*. As compared with the linguist, for example, the psychologist has done a very unsatisfactory job of defining his stimulus and response units on the purely descriptive level. Linguists agree among themselves quite well upon the identification of both the molecular units of messages (phonemes) and the molar units (morphemes), but it would be difficult to find two psychologists who, observing the same situation and activity, could agree on their identifications and classifications of stimuli and responses. I think we can take a hint from the linguist here, for, in actuality, he is dealing with a particular segment of the psychological domain, language, which is simultaneously both responses from one individual and stimuli for another. The linguist identifies two major levels of units: *phonemes* (which are meaningless in themselves but serve as basic discriminanda) and *morphemes* (which are meaningful units). Correspondingly, in behavior in general we have sensory and motor integrations ("phonemes") as the fine, molecular units of input and output as well as "situations" and "acts" ("morphemes"), but how are the latter to be defined?

Here again, methodologically, we can take a hint from the linguist. He classifies as "the same" all segments in messages which have the same *meaning* or significance in the code (thus the physically different endings of *walk-ed add-ed*, and *play-ed* in spoken language are all members of the same past-tense morpheme), without inquiring into *what* this meaning may be per se. Now, it seems that the representational mediation process in behavior makes possible a similar definition of molar units in psychology. I would say that all members of the class of physically different stimuli which are associated with a common representational process (e.g., have the same significance to the organism) are "the same" as molar situations; thus, a welcoming smile, "Hi!" "Good to see you again," and the like are functionally equivalent situations in the social interaction matrix. Similarly, I would say that all members of the class of physically different instrumental acts which are mediated by a common representational process (e.g., express the same intention for the organism) are "the same" as molar actions; thus, a tight-lipped glare, a muttered oath directed at someone, a clenched fist, and the like are functionally equivalent acts in the social interaction matrix.

This type of learning theory analysis

places a premium on developing methods of indexing representational states, e.g., measuring meaning, particularly in human organisms. Some research over the past few years has been directed chiefly along these lines (11). Dealing with language responses themselves—which, after all, are supposed to be "expressions of meaning"—attempts were made to discover a limited number of basic factors or dimensions along which meaningful reactions vary and hence can be measured. Several fairly large-scale factor analyses have already been done, and the same first three factors have appeared in each case: an *evaluative* factor, a *potency* factor, and an *activity* factor. A number of semantic measuring instruments have been devised, for specific problems on the basis of these factors, and these applications have been successful so far. It is assumed that these factors, derived in judgmental situations, are representative of at least some of the major ways in which representational mediation processes in human beings can vary.

ANALYSIS OF SOME BASIC SOCIAL MECHANISMS

It is a truism that the social behavior of human individuals is learned. The tremendous variation in personality and culture across the continents testifies to this fact. To the extent that the learning principles discovered by psychologists are valid, they should be operating in the complex matrix of social life, always conditioned, of course, by the particular situations of each culture. And if the analysis in the preceding few pages has been adequate, the most characteristic mode of learning at this level will be in terms of a two-stage process: learning of significances for signs and learning of instrumental acts which express intentions.

Assign learning—concepts, stereotypes, and symbols

In complex societies most signs representing categories and relationships are highly abstract, far removed from the earthy behaviors in which we presume meanings have their origins. We may refer to signs of this abstract sort as *assigns*, e.g., their meanings are literally assigned through repeated association with simpler signs having direct behavioral reference. An experiment recently completed will serve to illustrate this point (2). The problem was to demonstrate that the meaning acquired by a nonsense item, MEBLU, was predictable from the frequency of association of this item with simple adjectives. Subjects read a brief article, presumably taken from the *National Geographic Magazine*, in which they saw references to "the brawny Meblu" several times, "the excitable Meblu" twice, and "the friendly Meblu" only once; other groups of subjects encountered this hypothetical Meblu tribe with other adjectives and with different frequencies of association. In all cases the measured meaning of MEBLU shifted from neutral meaninglessness in a direction and to an extent predictable from the meanings of the common adjectives with which it had been associated.

Suppose now an individual is exposed repeatedly, and more or less exclusively, to the *Chicago Tribune*: He experiences the initially meaningless stimulus ALIEN in such contexts as "Aliens are *not* to be *trusted*," "Our national life is being *poisoned by* alien ideologies," and "We should *deport* these *dangerous* aliens." The resulting connotation of ALIEN is predictable from the meaningful contexts in which it appears. The important thing to consider is that most of the linguistic signs with which we deal in the mass media—Eisenhower, Fixed Farm Supports, the U. N., desegregation, and so on ad infinitum—acquire their meanings as assigns rather than through direct behavioral experience.

The notion of *stereotype* is that of a common meaning or significance attached to a class of people, not necessarily similar physically, this common significance serving to mediate a class of instrumental actions, variable with the total situation but all reflecting the common meaning of the class. Thus "stereotype" is a special case of the mediation model described in Figure 11. On the decoding side, the identification of a particular individual as a member of the stereotype (e.g., the evocation of the appropriate mediation process) usually depends upon certain combinations of cues—skin color, shape of nose, certain gestures, certain

dialectic deviations in speech, and so on—but the use of the appropriate linguistic label, in the newspaper, by someone else, is also sufficient to elicit the common mediation process. Since we all use stereotypes—indeed, categorizing of this sort is the essence of concept formation or "coding" of our world—the real problem concerns the *nature* of the mediation process or meaning. Either based upon direct behaviors as shown in Figure 11 (e.g., living in a race-conscious neighborhood where gang warfare along national lines occurs), or more often based upon assign learning through the words used by parents, friends, and the media we acquire such meanings—NEGROES may be conceived as friendly, kind, and submissive or as dangerous, cruel and aggressive depending upon such learning conditions. The classes of instrumental behaviors associated with stereotypes will necessarily reflect such significances, but will vary with the total situation in which the stereotype is experienced—Mr. A may hire a Negro for a menial job but avoid sitting next to him on the train, for example. The learning of such contextually dependent behaviors presumably depends upon the differential rewards and punishments delivered by those other individuals in the learner's environment (parents, superiors, etc.) who have prestige for him and with whom he identifies.

Abstract *symbols* play an important role in society. We have our flags, our ceremonial robes, and our briefcase in the hand. The distinguishing characteristic of the symbol is that it is a stimulus pattern which, already having its own specific representational process, is used to signify something else. For example, perception of a briefcase in a man's hand, as a visual pattern, specifically signifies a container for papers and the like; but above and beyond this specific denotation is its symbolic significance of "professional man." I suspect that we are dealing here with a rather direct process of transfer of meaning through association. Let's look at the hypothetical development of *money* as a symbol: The young child early learns to recognize nickels, dimes, and quarters, as well as paper bills, as objects having certain characteristics, round and rollable, hard and jingly, tearable and foldable and so on. But the nickel, let us say, is later used in exchange for an ice cream cone, and part of the ice cream cone significance (including its gratifying aspects) comes to occur in anticipatory fashion when nickels are held in the hand. If this were all that nickels were good for, they would presumably become a symbol of ice cream cones and nothing else. But later still another nickel is exchanged for candy, another for a ride on a bus, another for a pencil to write with, and even another to pay "protection" from the local bully. What do all these situations have in common? About the only thing is generalized gratification, getting something desired at the moment. As a result of such a learning sequence, one would expect money to acquire a very general "pleasure" and "security" significance, e.g., a very abstract symbol.

Imitation and identification

Another basic problem for social science is the transmission of culture (e.g., ways of perceiving and behaving in specified situations) from one generation to another. Deliberate training, of course, plays an important role, but much of the transmission is on a more informal, *imitative* basis. As Miller and Dollard (8) describe it, the dependent individual, not knowing the critical cues himself, learns to match his behavior to that of a model, an individual who does know the critical cues. For example, the older of two brothers in a family knows that heavy footsteps on the stairs signifies the arrival of father (often with a little present of candy or toys); he runs to the door; the little toddler in the family happens to imitate the older brother, responding to his running behavior by doing the same thing; both are rewarded with candy gumdrops. After a number of rewarded experiences of this kind, the younger individual will develop a generalized imitative set, a readiness to match his overt behavior with others in situations where he does not know the significance of the cues. Note that the prestige of the model here determines the probability of imitation, but this prestige in turn presumably depends upon rewarding associations.

But in many cases—most cases in adult behavior—what we observe is not simple

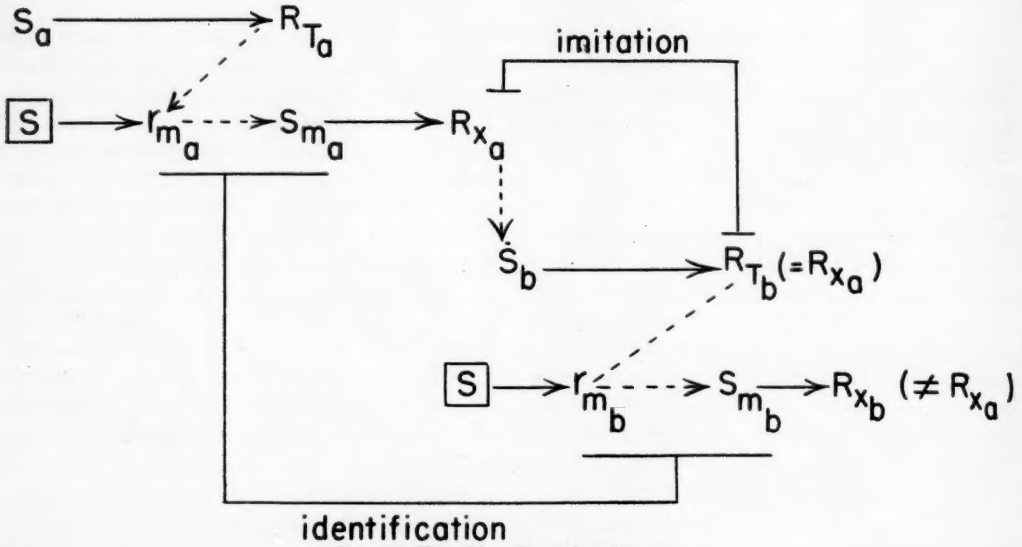


FIG. 13. The Sign-Learning Process.

imitation of the overt behavior but rather "appropriate" behavior in terms of the interpreted attitudes or meanings of the model. The teen-age son does not imitate his father's behavior toward girls. Not only has he had no opportunity to observe such behavior on his father's part, but times have changed and also courting customs. Rather he behaves "like his father might in this situation." When I see a person strike his thumb with a hammer, I do not behave like he does, but rather as I might behave in the same experienced situation, e.g., I "put myself in his shoes" and behave accordingly. We refer to this as identification of one person with another. But how does it operate?

Again we must use our two-stage model. Based on imitation of overt behaviors, e.g., of parent by child, the dependent individual acquires meanings of situations, people, activities, etc. that are similar to those of the model. Given such similarities in the significance of cues, the dependent individual will then behave in ways that are equivalent to those of the model, e.g., express the same intention, even though not identical and not based on direct imitation. Figure 13 illustrates the nature and development of such identifying behavior according to this type of theory. Subscript *a* refers to the model and subscript *b* to the dependent individual. Note that imitation refers to the identity of behavior to certain objects and identification to the similarity of meanings (representational processes) resulting. Thus, having imitated the authoritarian behavior of his father toward women in the family, the child develops a significance of women like that of his father, this condescending significance determining subsequent behaviors on his part to girls-in-general which, while not imitated from his father, are "like what he might do in such situations" (7). The broader import of such an analysis is that when some individuals in a society (followers) identify with another (leader), they do so on the basis of shared meanings for situations and then proceed to interpret situations as the model does and hence behave in ways equivalent to those of the model, e.g., quite literally "putting oneself in the place of another (interpretively) and behaving accordingly!"

Personality traits and culture traits

Allport (1) has defined a personality trait, in essence, as that which renders equivalent classes of situations and classes of behaviors for an individual. Thus, for a person with an "inferiority complex," any competitive situation signifies threat and he responds by avoidance, by "big talk," by compensating with some other activity in which he excels and so on, depending upon the particular context. Similarly, a person

with a "stingy" trait perceives situations like requests for loans from him, sharing the tip at a meal, and spending money for presents in similar ways and behaves toward these situations in equivalent ways, e.g., ways which have in common the avoidance of giving things up. Looking at Figure 11 again, it can be seen that this mediated generalization and transfer model applies here as well; a class of signs have acquired a common significance to the individual and this common significance mediates a class of equivalent instrumental actions.

Now, when the members of a society *share*, as a result of common experience and training, a mediational mechanism such as shown in Figure 11, I think we may speak of a *culture trait*. If, for example, members of a given society agree in perceiving spirits in natural objects like trees and lakes and agree in behaving appropriately toward these objects by ceremonial avoidance of some, giving offerings to others, talking about them with animistic terms, and so on, we have a *shared significance which renders equivalent a class of stimulus situations and a class of behaviors*, and thus, in my terms, a culture trait. At a more molecular level, all of the lexical items of a language can be thought of as culture traits. The concept "red" is a common cognitive reaction to a class of functionally equivalent stimuli, as is the concept "dog," the concept "happy" and so on. I do not know whether this conception of a culture trait is consistent with sociological theory or not, but it does offer a hand-hold for analysis of the formation of such shared significances and behaviors.

SUMMARY

Starting quite frankly from the notion of the hierarchy of the sciences, this paper has tried to show how psychology, as a sort of Janus-faced discipline between the biological and social sciences, both implies certain neurophysiological conceptions and is implied by certain social science conceptions. The basic principles of simple S-R psychology were described and then shown to be insufficient (but not necessarily invalid) for an analysis of characteristically human behavior. A two-stage, mediational model was then suggested, a model which emphasizes both the decoding of situational significances and the encoding of intentions instrumentally. The implications of such a model for the observed flexibility of human behavior, for a distinction between "primary" and "secondary" learning, and for the problem of psychological units were drawn. Finally, a very superficial and sketchy application of this type of psychological theory to a few socially relevant mechanisms, like stereotyping, identification, and culture traits, was made by way of illustration.

REFERENCES

1. Allport, G. W. *Personality: a psychological interpretation*. New York: Holt, 1937.
2. Dodge, J. Predicting the meanings of assigns from the measured meanings of signs with which they are associated. Unpublished doctoral thesis, University of Illinois, 1955.
3. Guthrie, E. R. *The psychology of learning*. New York: Harper, 1935.
4. Hebb, D. O. *The organization of behavior: a neurophysiological theory*. New York: Wiley, 1939.
5. Hull, C. L. Knowledge and purpose as habit mechanisms. *Psychol. Rev.*, 1930, 37, 511–525.
6. Hull, C. L. *Principles of behavior: an introduction to behavior theory*. New York: Appleton-Century-Crofts, 1943.
7. Lazowick, L. M. On the nature of identification. *J. abnorm. soc. Psychol.*, 1955, 51, 175–183.
8. Miller, N. E., & Dollard, J. *Social learning and imitation*. New Haven: Yale Univ. Press, 1941.
9. Osgood, C. E. *Method and theory in experimental psychology*. New York: Oxford Univ. Press, 1953.
10. Osgood, C. E. A behavioristic analysis of perception and meaning as cognitive phenomena. *Symposium on cognition*. University of Colorado, 1955.
11. Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana: Univ. of Illinois Press, in press.
12. Pavlov, I. P. *Conditioned reflexes*. London: Oxford Univ. Press, 1927.
13. Skinner, B. F. *The behavior of organisms: an experimental analysis*. New York: Appleton-Century-Crofts, 1938.
14. Thorndike, E. L. *Animal intelligence: experimental studies*. New York: Macmillan, 1911.
15. Tolman, E. C. *Purposive behavior in animals and man*. New York: Appleton-Century-Crofts, 1932.

(Manuscript received December 15, 1955)

A CALCULUS FOR ETHICS. A THEORY OF THE STRUCTURE OF VALUE. PART II

by Nicholas M. Smith, Jr.

Operations Research Office, The Johns Hopkins University

This is the second of two articles[1] dealing with a theory of values based upon a stochastic model of a system. Several ethical systems—utilitarianism, casuistry, moral idealism, Hegelianism, and pragmatism—are discussed in terms of this theory and finally an underlying and unifying rational principle is proposed, the principle of *analogical conformity*. It asserts that all rational forms, including many laws of various natural and social sciences as well as ethical principles are homomorphic to a general form. This form is a generalization of probability theory.

I. A DESCRIPTION OF SOME ETHICAL SYSTEMS IN THE TERMINOLOGY OF VALUE THEORY

It is the object of this section to describe in the technical terminology of value theory particular ethical systems. Owing to limitations of space, time, and degree of development of the theory, these illustrations are necessarily somewhat superficial. Five different ethical systems will be discussed and then some attempt made to correlate the ethical systems with value theory. In order to present the ethical systems in a most condensed form the outline of various systems as appearing in Leys' *Ethics for Policy Decisions* will be quoted (5). In this outline Dean Leys condenses a particular ethical system into a set of questions from which the individual decision-maker is to obtain answers in order to solve his decision problem.

A. Utilitarianism

This ethical system is concerned with the relationship of proposed action to the concept of utility and it most nearly resembles value theory applied to an elementary model. According to Leys the following questions are asked:

Utilitarianism (Bentham)

1. What are the probable consequences of alternative proposals?
2. What policy will result in the greatest possible happiness of the greatest number?
3. How do the alternatives compare in the intensity, duration, certainty, propinquity, fecundity, purity, and extent of pleasures and pains?
4. What proposed actions are protected from a scrutiny of probable consequences by sacrosanct phrases?
5. What is the factual evidence for assertions about benefits and disadvantages?

The first question concerns the relationship between the probability of outcome and action taken. The second question might be described as the elementary criterion of decision in which values of ultimate states are weighted in proportion to "the greatest happiness for the greatest number." Which action is associated with the maximum value? The third question relates to the postulates of value of the ultimate consequences. It implies that the good of value is time-dependent and inquires into the intensity and duration and time of initiation of the pleasure states. Questions 4 and 5 refer to the validity of the arguments concerning the probabilities of transitions. Although the pleasures referred to may be transitory, the ultimate good for the greatest number must be estimated in a manner not too clearly defined. If concrescence occurs in utilitarian ethics, it must occur spontaneously in the masses; there seems to be no explicit provision for the evolution of utilitarian ethics except as the consensus of the masses fluctuates.

[1] Part 1 of these two articles appeared in the April 1956 issue of *Behavioral Science*, pp. 111–142.

B. Casuistry

Casuistry is the ethical system justifying the legal profession. "Casuistry" is a technical word referring to the case-history method of analysis. Although this word has an undesirable connotation of sophistry, it is in the technical sense of case history that the term casuistry is used here. According to Dean Leys the casuist asks the following questions:

Casuistry

1. What are the authoritative rules and precedents, the agreements and accepted practices? What would be ———'s attitude?
2. What citations are lacking in authority or acceptance?
3. What citations are irrelevant to the case in hand? Which are irrelevant, because they were not intended to apply, because they were not intended literally, because of expressed or unexpressed qualifications, because they were counsels rather than commands, because their meaning can be distinguished from the case at hand, because an application to the present case would not serve the original purpose of the authoritative rule?

The basis for casuistic ethics in value theory is to be found in the fundamental theorem of value, (Part I, Section III, A). This theorem states that if the fundamental value postulates are held constant, and if the transition probabilities are independent of the absolute time—i.e., if they depend only on the difference in time –and if furthermore the system remains constant—i.e., has no new states added nor states removed—then the value associated with the state of the system is a constant quantity. This means that whenever a system returns to a state in which it has existed previously, the same values prevail in the present situation as prevailed in the historical situation. The decision criteria are reversible. As stated, decisions are made on the basis of maximization of expected values. On the other hand, historical decisions imply values that existed at the time. Hence, if facing a particular situation in the present an individual can search history and find a case that was identical to his own, i.e., find decisions that were made when the system existed in the same state as he now faces, then he may assume that the values pertaining to his present situation are the same as those that were faced in the historical situation. Decisions that led to desirable results are studied; these are said to be authoritative and to imply values that existed at that time, and, by implication, serve as a precedent for decision in the present. *Casuistry is the method of experimental ethics.* In Dean Leys' outline, Question 1 searches for authoritative cases. Question 2 is concerned with the elimination of nonauthoritative cases. Question 3 is essentially determining if the conditions given in the fundamental theorem apply; namely, is the historical case identical with the present case? Is there irrelevant information that camouflages the true situation? Although not included in the above set of questions, another question the casuist should ask is whether the fundamental value postulates have changed in the interim.

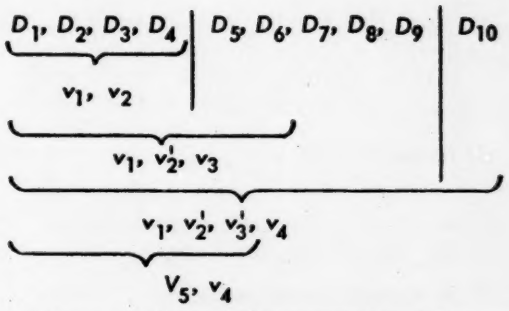


FIG. 1. A Schematic Representation of the Evolution of a Value System by Concrescence.

D_1, D_2, etc. are decision problems. The v_1, v_2, etc. are the value postulates consistent with the decisions. The vertical lines are conflicts (inconsistencies) appearing when a new decision is faced. A new value system is reached by concrescence; the newest value usually being the one adjusted. Occasionally several elemental values are consolidated into one value (moral) construct, (V_5).

Figure 1 is a diagram of the evolution of a value system by concrescence.[2] This representation of the evolution of the value system is also a description of an experimental method for determining the values existing in another system. The greatest

[2] A more detailed analysis of the evolution of a value system is given in Part I (*Behavioral Science*, July 1956, p. 138 ff.).

problem the casuist has in applying his method to decision problems in the present is that of assuring himself that the case studied is in fact identical to his own.

C. Moral Idealism

In Dean Leys' summary the moral idealists ask the following questions in facing his decision problems:

Moral Idealism (Plato and Kant)

1. Can you define what you approve? Can you bring various approved practices under a general rule? Does the definition or generalization state what you always approve? Does it cover only those practices that you approve?
2. If there is a conflict of principles, can you find a more abstract statement, a "third principle," which will reconcile the conflicting principles?
3. What is your scale of values? Are you putting first things first?
4. What would you approve in an ideal community?
5. How can you be worthy of happiness?
6. Can you will that the maxim of your action should become the universal law?
7. Are you treating humanity as an end and not merely as a means?
8. Are you legislating for yourself, that is, laying down a rule that you recognize as right?

Many of these questions concern the existence of a desired goal, and the willingness of the individual that his desired goal will fit in with the concept of an ideal community; that his rules for action, his ethical code, be adopted by the remainder of the community. In general, idealism requires some reference to absolute values.

In particular, look at Question 2. In the presence of the conflict of principles, can a more abstract statement, an all-inclusive third principle be found that will reconcile the conflicting principles? This is, of course, the process of concrescence.

D. Hegelian ethics

Dean Leys sums up the questions asked by Hegel as follows:

The Historical "Logic" of Hegel

1. What are the main historical trends of institutions? What is the direction of national development?
2. How are our partisan causes related to this larger picture? How may my interests and ideas be reconciled with the interests and ideas of the opposition in the historical evolution of the nation? What is my station and its duties? To what can we all be loyal? In what can we all participate?
3. What synthesis has overcome the opposition of thesis and antithesis?

Hegel seems to be primarily concerned with the trends of institutions, the effect of actions on these institutions, and the reconciliation of individual actions with the historical evolution of these institutions. It appears that in the terminology of this discussion Hegel is concerned with the values relevant to the largest and most inclusive systems in the hierarchy of systems to which an individual belongs. That is, he is concerned with *total optimization* and the avoidance of suboptimization. In particular, in Question 3, he is concerned with the process of concrescence as a means for the resolution of conflict.

E. Pragmatism

Pragmatism, or as Dean Leys calls it, "Dewey's Instrumental Thinking," is the ethical system most concerned with dynamic evolution of values. Dean Leys' summarizing questions are as follows:

Pragmatism (Dewey)

1. What is the problematic situation that gave rise to deliberation?
2. What will satisfactorily terminate deliberation, that is, relieve the conflicts and tensions of the situation? Does the proposed solution anticipate consequences in the larger environment as well as the immediate situation?
3. Are our ends changing as well as the means of achieving ends? Does this evolution of ends and interests affect the merits of proposed solutions?

After defining the problem and clearly isolating the tensions and conflicts involved, Dewey asks what action will relieve the conflicts? In value theory the resolution of conflict has been considered an absolute goal toward which value systems evolve. In the latter part of Question 2, concern is expressed that the consequences of the action solve not only the immediate tension, but also be a solution of any anticipated tensions in the larger environment as inferred from the history of the system. This is precisely the condition that the concrescence process in relieving one conflict not introduce another. A value system must be con-

sistent not only with the present situation but with the past history of the system and the anticipated history of the system. The last question refers to the possible evolution of the value system and evolution of the values placed upon the ultimate states and ultimate goals. Dewey may be considered the relativist in ethics; his values evolve in such a direction as to relieve conflicts and tensions. There is no concept of absolute values. In Section III, the question of absolute vs. relative values will be considered in more detail.

II. EMPIRICAL METHODS

It has already been pointed out that the fundamental theorem, together with the reversal of the criterion of decision is the basis for empirical methodology based upon case history studies, i.e., of casuistry. It will be instructive to take a particular empirical methodology for the practical determination of values in certain areas and point out how this fits into the structure of value theory. The example chosen will be that described by Churchman and Ackoff as "An Approximate Measure of Value" (1, 2, 3). This illustration of the method as described by the authors will also lead very naturally into the next section where a discussion of the epistemological problem in value theory is made.

The problem as posed by Churchman and Ackoff concerns the establishment of basic values of customers for given commodities. They make the following assumptions:

A-1: For every outcome O_j, there corresponds a real nonnegative number V_j, to be interpreted as the *true* [italics mine] importance of O_j.

A-2: If O_j is more important than O_k, then $V_j > V_k$, and if O_j and O_k are equally important, then $V_j = V_k$.

A-3: If V_j and V_k correspond to O_j and O_k respectively, then $V_j + V_k$ corresponds to the combined outcome O_j and O_k.

Assumption A-3 is basically the assumption that the values of things are additive. It has already been demonstrated that the values of things are additive only for marginal situations; i.e., for situations in which actions do not affect importantly the ultimate goals of the system. Various corollaries to the assumption of additivity are mentioned by the authors. It is sufficient to point out here that the V_i's are conceived of as the true absolute values of which estimates v_i are to be made. The method proposed provides the technique, presumably, for narrowing down the error between the absolute values V_j and its estimate v_j. In this method an individual is subjected to two tests, each of which contributes information concerning the importance of the outcomes with respect to his own value system. In the first test, the individual assigns names and tentative quantities to values involved along the scale provided for him by the experimenter. Next, he is confronted with certain questions concerning relative preferability of certain random combinations of the outcomes. That is, he is asked to judge essentially what decision he would make if he were placed in a situation where he would have to express his preference. The subject provides the factorization of the values and gives them subjective quantization. The values are ranked in the order of decreasing importance and tentative numbers assigned to them. Next, on the basis of the relative preferences of various combinations of outcomes the numbers are adjusted for consistency. For convenience their sum is normalized to unity.

The reader will recognize this methodology as the casuistic method that pertains to marginal situations; i.e., the situations for which Assumption A-3 may be considered valid. Although the limitations of the assumption of linear additivity are very stringent, a great deal of very useful analysis and research is possible under these conditions.

It is important to distinguish between a true casuistic procedure and a pseudocasuistic procedure. A procedure that requires no actual decision to be made but merely requires a statement of opinion as to what decision might be made in a given situation is referred to as a pseudocasuistic procedure. A true casuistic procedure does not accept statements of intended decision, but can only accept actual decisions that have been made. For instance, it is impossible, or certainly very difficult, for an

individual to imagine his actual feelings in a situation greatly different from the one in which he is presently placed. A man might be asked how he would act in a situation of extreme danger. His belief as to what conduct he would follow may or may not be very well correlated with his actual response. The situation that is presented to the individual for the estimated decision must be one that departs very little from the situation in which the individual is usually placed. As the method of Churchman and Ackoff has been applied, this condition is valid. Housewives are asked what color of package they prefer, etc., all involving hypothetical situations not very far removed from their everyday life.

Notice the similarity between the development of estimated values by the method of Churchman and Ackoff and the description of the evolution of value systems. Case examples—that is, possible decisions—are studied and a set of trial values is arbitrarily selected in such a manner that the ordering of these values is consistent with the trial decision studied. This process is identical to the process that has been described here as the arbitrary, intuitive postulation of values. Furthermore, the condition of consistency (no conflicts present) is necessarily imposed by the value estimators. Whenever new data (i.e., decisions) present themselves it may be necessary for the experimenter to readjust his value estimates to restore consistency. Such a readjustment, or perhaps even a refactorization of his values, constitutes what has been described here as the concrescence process.

III. THE EPISTEMOLOGICAL PROBLEM IN VALUE THEORY[3]

The discussion by Churchman and Ackoff and the concept of the true and absolute values employed in their paper suggest two diametrically opposed viewpoints as to the nature of ultimate values. One has from their viewpoint a concept of absolute values, the discovery of which becomes the prime objective of the experimentalist. Or, one may adopt the viewpoint that ultimate values are essentially relative, that they are constructs of human rationalization. This difference of viewpoint is a reoccurrence of an ancient dispute as to the reality of existence of the external world versus its existence purely as mental constructs of the subject. This entails the epistemological problem.

Each viewpoint has its own particular pitfall into which the unwary may fall by misuse of his methods.

The rational absolutist acknowledges that the values he quotes are but estimates of the true absolute values. His chief problem then becomes one of estimating the difference between his estimates and the true absolute values. He has no way of estimating his absolute error and must constantly acknowledge that the values he employs in decision processes are but estimates of the true values. His chief pitfall is that he is apt to forget that the values he is using are but estimates and are, therefore, subject to error. When he does forget, in the belief that some of his values are the true and absolute ones, he becomes trapped in his own methodology. Absolute values will not permit concrescence, the growth and evolution of a value system. It is no longer possible for the individual who believes he is in possession of the absolute values continuously to re-evaluate his value estimates.

The relativist, on the other hand, has his own problems and potential pitfalls. In the viewpoint of a relativist, values of ultimate states and goals are but postulates made by him in an intuitive manner and thus are not unique, and, in a sense, are arbitrary. His chief problem becomes one of judging the "goodness" of his set of arbitrary values with respect to any other set that might be used. The chief pitfall into which he may fall is that he may too readily admit that his value system is arbitrary and dismiss and reconstruct it purely as a matter of expediency in solving the particular problem at hand, relieving the local tension without due consideration of the effects of this reconstruction on other situations. The absolutist may become too rigid; the relativist too pliable.

[3] The material of this section originally appeared in the Comments by the author appended to a paper by Churchman and Ackoff (Reference 3).

It is not to the possible misuse of value methodologies to which reference is made, however, but to their consistent and logical implications. Consider an external observer noting the actions of a particular individual and attempting to determine by these actions whether this individual is an absolutist or a relativist. If the individual under observation is a relativist, he is observed to make postulates of values of ultimate states, to make decisions on the basis of these values, etc., until a tension created by conflict is reached. Then through the process of concrescence, he readjusts his values to relieve the tension. Where two or more value systems are possible that relieve the tension of conflict, the criterion of simplicity is invoked to select that system involving the simplest set of postulates; i.e., simple in terms of the number and nature of the fundamental value postulates but not necessarily with respect to the complexity of the logical operations involved. Thus, the relativist is resorting paradoxically to an absolute criterion of rational virtue, the criterion of consistency (nonconflict) and of simplicity.

Now, on the other hand, assume that the individual is an absolutist. He is seeking to discover or approximate an estimate of the absolute value. He is observed to select a set of trial values *arbitrarily* in such a manner that their ordering is consistent with the decisions observed. Whenever inconsistency occurs as new decisions are considered the arbitrarily selected trial values are readjusted. The absolutist continually keeps in mind that his set of trial values are but estimates of the real absolute values. He resorts, paradoxically, to a relative methodology. To the external observer the operations of these two individuals are indistinguishable. It can be concluded, therefore, that there is no more sense to the epistemological argument in value theory than there is in physical theory. An individual may select either viewpoint, as he prefers.

Since the absolutist is never able to identify values as absolute, these absolute values are never involved in his operation. Since the possibility of discrepancy between the estimates of values and the true absolute values always exist, I personally see no need for the postulation of existence of absolute values. It must be acknowledged, however, that these viewpoints are purely a matter of personal preference.

IV. THE STRUCTURE OF MORALE

In this section a factorization of the components of morale in the terminology of value theory will be attempted. These components will be primarily ethical in nature. This is presented in an attempt to illustrate the potential clarification of a complex construct through factorization into value terms.

1. In the first place, it is observed that the subject of morale enters into the relationship between groups and subgroups, individuals within organizations, organizations within communities, communities within nations, etc.

The first factor selected then is a *hierarchy of groups*. An individual may belong simultaneously to many of these groups—his self-system, a group of close friends, and various organizational groups, partially overlapping.

2. Next it is observed that each group forms a system, and *associated with each system is a set of values* corresponding to the expressed or unexpressed goals or pleasures of that system.

3. It has been observed that whenever one set of states exist simultaneously in various groups, conflicts will sooner or later develop. These conflicts are resolved by various schisms or by formation of a super group whose values are considered tantamount. It is recognized that one of the components of morale has to do with a ranking of the related group values. This ranking produces the conditions for *a set of vector values*. This ranking of values will result, under stress, in the repudiation of the values of some subsystems before that of others. A stress in a situation is brought about by a state of conflict.

4. A group has leadership. This leadership is delegated to the leader from subgroups within it by means of *a transfer of decision prerogatives*. The leader acts for the

group within certain matters whose bounds are defined as areas of authority.

5. Two things now result. One is the feed-up of values from the constituents to the leadership, a sense of loyalty to leadership, that is, *a rigidity on the part of the subsystems with respect to the values established by leadership*, or established by concrescence of the whole. In the other direction:

6. The constituents demand evidence of a *sense of responsibility by leadership to the "rights" of the constituent groups*. The authority of leadership and the rights of constituent groups are fundamental postulates of government. Our own Declaration of Independence contains such postulates:

> We hold these truths to be self-evident, [we establish these postulates] that all men are created equal; that they are endowed by their Creator with certain unalienable rights; that among these, are life, liberty, and the pursuit of happiness [*postulates of individual rights*]. That, to secure these rights, governments are instituted among men, deriving their just powers from the consent of the governed [*an acknowledgment that the true power of leadership is derived from the transfer of prerogatives from constituents*].

7. Both groups, leadership and constituents, must be continually informed of the states of stress to each other. The systems are stressed internally as well as externally. Leadership must be continually aware of the stresses to which it is placing its constituents. Conversely, the constituents must be aware continuously of the measures taken by leadership to relieve their stresses. This makes very necessary and important to the state of morale a *two-way communication system*.[4]

8. Lastly, it is recognized that the degree of rigidity of the constituents to the value system of the group is increased by confidence resulting from a *successful history* by the group *of withstanding stresses*. This is the value of tradition to an organization. Ceremonials serve to pass on to new individuals the history of successes.

[4] The importance of this factor has been pointed out by Dr. David McKay Rioch, Director, Neuropsychiatric Division, Walter Reed Army Institute of Research, Walter Reed Army Medical Center, Washington 12, D. C.

The values of a system are thus oriented in three distinct directions: within, to the base or constituent subsystems and the constraints their value stresses place upon the system; to the self-system values of survival, etc.; and lastly, to the values of the super-system to which the system itself is subordinate. The three components of personality, the id, the ego, and the super-ego are thus recognizable in this general value structure.

V. THE SUPREME POSTULATE

It would be amiss in a paper attempting this degree of comprehensiveness not to try placing the Supreme Postulates in the formalization of value theory. These postulates are concerned with the concept of the Self, the doctrine of the indestructibility of the human Soul, and the concept of God. In this Section it is unavoidable that we tread on metaphysical grounds.

Of the three postulates mentioned above the concept of the Self is believed to be most fundamental. Somewhere in the growth of the individual from conception through birth and early awareness, the concept of its own individuality must emerge. It seems apparent that this concept must arise very early in this development. It is even believable that the notion of the Self originates very close to the birth of the child and in some aspects before birth. However the development of the Ego must be a continuing process. It is possible that the first concept of Self is universal, i.e., includes all factors of the individual's experience both "internal" and "external" to him. Early awareness may be unable to differentiate between the Self and the non-self.

One must either assume that the individual is so constituted in his structure that first thoughts occur in terms of an over-all concrescence of all the experience factors; or, one must assume that this over-all concrescence occurs after the first awareness. The latter seems preferable, since during the first history of early awareness this concrescence must have to be adjusted many times as new experiences unfold. Imperfections in the Self concrescence may even occur.

Descartes is famous for his saying "Cogito, ergo sum:" "I think, therefore I am" which he regarded as irrefutable evidence of the existence of the Self. This statement bears analysis. "I think" presupposes the "I" from which it is later concluded that "I exist." In what sense is this statement irrefutable? It is the nature of the approach to values and concepts in this presentation to regard all constructs as postulatory in character. This viewpoint may be extended to include these three fundamental concepts. In a sense every concrescence determines an Ego, and the Self of a particular system is defined or created by this concrescence. This is particularly noticeable in military organizations, for example. Somewhere in the development of a system there becomes an awareness of oneness or singleness of purpose—that is, of a common set of values brought about by concrescence at that level. This awareness in the military units is called *esprit de corps*. The establishment of a strong *esprit de corps* is one of the principal objectives of leadership and implies a high state of morale. A similar process must happen within the individual. He may consider the Self to be universal in the beginning and only as more experience indicates that he has control of only part of the factors in his universe is the concept of the Self differentiated from other external factors in the universe. The postulate of the Self may be considered a point of departure; a defining construct.

As an analogy, if one were in a new undeveloped land without any instruments other than native material which could be picked up to make marks, he might first make a mark on the ground or drive a stake and state that that stake would be the primary point of reference for all subsequent measurements. He might then drive a second stake an arbitrary distance away and in an arbitrary direction and state that the line and direction so determined define a fundamental reference direction and a fundamental unit of measurement. These acts would then lay down the necessary framework for a local survey of the new territory. It is in this sense that one may regard the construct of Self. "I think, therefore I am," rather than being a fundamentally irrefutable proof of the existence of the Self has the nature of a tautology.

The emergence of the Self from the libido must occur then at some time as a fundamental—perhaps universal—concrescence. This process can be imperfect. One or more concrescences may occur with overlapping areas. Schisms may occur which permit two postulatory frameworks to develop almost independently side by side, resulting eventually in the phenomena of multiple personality.

The phenomena of existence or reality in this presentation have been taken to be measured by the workability of a construct in the decision making process (i.e., nonconflict). The concept of the Self being in a sense a fundamental point of departure then serves continually as a usable construct in all the postulatory systems which follow. Hence it must have a highest degree of reality.

The doctrine of the Immortal Soul cannot have the same tautological nature as the concept of the Self. It may originate from the very strength of the Self concept. If, as suggested above, the original concept of the Self is universal, subsequent experience can be expected to shrink this concept to a lesser and lesser role with respect to the universe of experience. These properties of the Self must be surrendered with reluctance, as exhibited by the fierce egocentricity of young children. We have postulated that one of the ultimate fundamental goals of any system is survival; that the concept of survival implies an assumption of immortality. Sooner or later the individual must become aware of the mortal nature of his own physical self.

The doctrine of immortality of the Soul is itself a postulate that permits acceptance of the mortal role and continued rational action in the face of inevitable destruction. In the viewpoint of value theory the question whether the Soul really exists must again be referred to the criterion of workability of the concept. If it is a construct which permits a continued sequence of decisions without itself introducing conflict, and the subsequent tensions produced by

these conflicts, it may be considered to have reality. Thus it is a reality which will be judged on the same scale in which one judges other moral and materialistic constructs.

A second description of the development of the construct of the Immortal Soul arises from inferences that the Self persists in a rapidly changing World. Throughout his life, a person observes the short time-span of things, political institutions, and even some ideas. But there is an experience which endures, despite changes in the external world. This is his awareness of his existence as a Self. To be sure, that Self may change with each concrescence, but as an entity, it seems to endure. The quest for a rational substructure to the world leads, in a natural way, to the postulation of a part of the Self which is Immortal: the Soul.

The concept of God is more complex since throughout the ages this construct has been employed for many different purposes. Man's knowledge of external world is imperfect. Many phenomena have been observed that cannot be related to other phenomena he connects by rational explanation. A supernatural being may be postulated which makes this connection. These early Deities were employed to "explain" the imponderable. At one time in some cultures every bit of imponderable nature had its own Deities: Gods of the Winds, Seas, the Sun and Stars, of Thunder and Lightning; of the Dark and Unknown.

A concept of a Deity has been employed in various times in history as a source of authority for government (such as divine right of kings), as the source of social control, and even as a source of hygiene and dietary regulation. The concept of God is also employed to make more consistent the doctrine of Immortality of the Soul. As more and more has been understood of nature, properties of the deity have changed. The concept of many gods has, by concrescence, reduced to but One. As one sees the doctrine of the Immortal Soul as one solution for the acceptance of the mortal role, so the concept of a Deity reinforces the doctrine of the Soul by adding consistency and simplicity in the over-all postulatory framework.

Is God real? Again one must impose the concept of reality herein considered. The construct of God is real in proportion to the degree of consistency and lack of conflict it permits in rational decision process. Other properties afforded the Deity again must be interpreted in terms of the consistent picture developed.

VI. THE PRINCIPLE OF ANALOGICAL CONFORMITY

In this section a principle relating to the formalistic aspects of reasoning will be enunciated and arguments presented for its validity. These arguments will be incomplete. A great deal of investigation remains to be done to establish the principle and demonstrate its potential usefulness. The *principle of analogical conformity*, as it is called,[5] is suggested by the formalistic congruency in the evolution of values and the scientific method and in the theories of various sciences, such as physics, economics, psychology, etc.

A. Statement of the principle

The most general statement of this principle is that *there exists a single rational form and that the form of any particular science* or reasoning *is homomorphic to this general form*. This general form is thought to be a more general form of probability theory. A more restrictive definition of this principle can be made. Certain similarities in the form of probability-value theory with the form of quantum mechanics in physics will be observed below. In quantum mechanics certain mathematical operators[6] acting on the scalar product of the wave

[5] This name was suggested to me by Dr. George E. Harmse, formerly of the Operations Research Office.

[6] An "operator" is a symbolic manipulation of mathematical quantities. Multiplication, addition, integration, differentiation, etc., are mathematical operations symbolized by x, $+$, $\int$, $\partial/\partial x$, etc. Operators may be manipulated separately from the operand as, e.g.,

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\phi,$$

where each operator is applied successively to the function ϕ.

function with its adjoint—that is in this particular case, its complex conjugate—produce expected values of measurable observables. In the restricted statement of the principle of analogical conformity, it is stated that the *same operators that produce observables in the physical system when operating on the scalar product of wave functions produce in the general system quantities that behave in that system in an analogous manner.*

For example, the operator,[7] $\partial/\partial x_i$ (the partial derivative with respect to x_i) where x_i is a coordinate of a physical system applied to the scalar product[8] of the wave functions, yields an expected measurement of momentum, $\bar{p}_i$, in the direction of the i^{th} axis:

$$\bar{p}_i = \left(\psi^* \cdot \frac{h}{2\pi i}\frac{\partial}{\partial x_i}\psi\right). \qquad (1)$$

In general, the operator O_i

$$\bar{O}_i = (\psi^* \cdot O_i \psi) \qquad [2]$$

produces an expected measurement of the corresponding observable. The same operator may be applied in a similar manner to the scalar product of the adjoint probability with the forward probability; e.g.,

$$\left(P\dagger \cdot \frac{\partial}{\partial x_i} P\right) \rightarrow (\bar{p}_i), \qquad [3]$$

where the dagger (†) indicates the backward equation. According to the principle of analogical conformity, the quantity so produced will be an observable with respect to the general system, which will have a mathematical form with respect to that general system in congruity with the form of momentum of the quantum mechanical system. The symbol ($\bar{p}_i$) of Equation 3 represents this logical congruency. Should the principle of analogical conformity be established, its potential usefulness will lie in the fact that the theoretical advancement in any one science can serve as a guide for theoretical advancement in any other. This does not mean that the congruity can be applied blindly, since the form of each science is taken to be a special case of a more general form. For example, it can be expected that the laws of physics will have their counterparts in game theory, economics, psychology, and other sciences. The principle of analogical conformity implies that these laws, related as they are to certain operators in the physical theory, will be produced by these same operators in the other sciences.

It is as if the great laws of science such as those of the conservation of momentum, the conservation of energy, the second law of thermodynamics, etc., are the inevitable product of the form of the rational process. The problem of science becomes one of determining the substance of experiment and observation that enables the formation of the constructs that lead to these inevitable formulations.

B. Evidences and plausibilities relating to the principle

The arguments to follow are somewhat sketchy and imprecise. A great amount of time, meticulous work, and moments of insight are required for their rigorous demonstration. It is hoped that at the very least they will be provocative in character. They are published here in the hope that investigation by others may be motivated.

1. Quantum-mechanical analogies. Certain formalistic analogies between probability theory and quantum-mechanical theory are apparent. In quantum-mechanical theory the scalar product of the wave with its complex conjugate is postulated to produce a probability function. The wave function itself is called a probability amplitude function. It is also postulated that the ψ functions are superposable; i.e., if ψ_1 is a wave function appropriate to the description of one system, say an electron in motion through slit 1 and ψ_2 is that function appro-

[7] When this operator is applied to a function defined in a discrete space it will be understood that the difference-operator, $\Delta/\Delta x_i$, is meant.

[8] In the notation used here, the parentheses $(\phi_1 \cdot \phi_2)$ indicates the integral $\int_\tau \phi_1\phi_2\, dt$ where τ is the phase space appropriate to the integral if the functions ϕ are continuous; or indicates the summation

$$\sum_k \phi_1(k)\phi_2(k)$$

if the functions ϕ are defined on the discrete denumerable set k.

priate to the description of an alternative route through a second slit, the wave function describing the combined event is given by the sum of the wave functions of the separate events:

$$\psi_{12} = \psi_1 + \psi_2 . \qquad [4]$$

The probability associated with the combined events is then

$$(\psi^*_{12} \cdot \psi_{12}).$$

For the moment, consider that the wave function, ψ, is in fact a probability function and that its complex conjugate is in fact an importance function. It will be necessary, of course, to assign a meaning to complex probability and value. It can be shown (4) that the forward probability function expressed in Equation 4 in Part I may be rewritten in the form:

$$\frac{\delta P}{\partial t} = PM, \qquad [5]$$

where M is a matrix operator that acts only on the space coordinates of the probability function. The backward or importance equation may also be expressed in this form, as in Equation 4:

$$\frac{\partial P\dagger}{\partial t} = MP\dagger. \qquad [6]$$

The probability function in Equation 5, for example, may be separated into two functions, one a function of space alone, and one of time alone.

$$P = S \cdot T, \qquad [7]$$

where S represents the function of the space coordinates only, and T a function of the time coordinates only. This separation of variables produces an arbitrary eigenvalue ω:

$$\frac{\dot{T}}{T} = \frac{SM}{S} = \omega \qquad [8a]$$

and it is seen that the solution of the time equation is of the form

$$e^{\omega t}, \qquad \omega \leq 0,$$

where ω is restricted to nonpositive magnitudes in attrition systems. Thus it may be concluded that a real state probability is to be associated with a superposition of states that decay exponentially, that is, with transient states. On the other hand, if the probability function had been imaginary, the eigenvalue would be imaginary, $i\kappa$.

$$\frac{\dot{T}}{T} = \frac{SM}{S} = i\kappa. \qquad [8b]$$

The solutions to the time equation in this case would be $e^{i\kappa t}$, and the over-all solution would be a superposition of such oscillating states. Thus, it is indicated that an imaginary probability function, if any exists, would be appropriate to the description of oscillating, or as is said, stationary states. Similarly, it is indicated that imaginary value is to be associated with the net increase of the amplitude of the forward stationary oscillating probability. These statements all apply to conditions under which all the elements of the matrix operator are real. There remains to be discussed an interpretation of an imaginary component of an elemental transition probability. Some indication of such an interpretation will be produced in the discussion on the special theory of relativity to follow.

Nevertheless, assuming for the moment a use for a complex probability notation and for equations of type 5 and 6, it is seen that the solutions would naturally be superposable because of the linearity of the differential equation. In real probability theory this results that the probabilities associated with mutually exclusive events are additive; in quantum mechanics the same rule applies but the imaginary probability has been interpreted as a "probability amplitude." Next, it would also be natural to associate an expected importance with the scalar product of the probability of the complex probability function with its adjoint. Since the quantum mechanical equation is self-adjoint in complex notation, the complex conjugate is the adjoint function of the forward wave equation and may be interpreted as a value function. Thus it appears that the scalar product of probability with value is analogous in value theory to the scalar product of the wave function with its complex conjugate in quantum mechanics, both being *expectations*.

TABLE 1

ANALOGY OF OBSERVABLES PRODUCED BY OPERATORS IN Q MECHANICS AND IN VALUE THEORY

Operator	Expected magnitude of the observable in	
	Q Mechanics	Value Theory
x_i	$\bar{x}_i$ i^{th} coordinate of position	$\bar{x}_i$; useful inventory of i^{th} commodity
$\partial/\partial x_i$ or $\Delta/\Delta x_i$	$\bar{p}_i$ component of momentum in direction of i^{th} axis	$-\bar{r}_i$, avg value of thing of type i

$$(P\dagger \cdot P) \sim (\psi^* \cdot \psi). \qquad [9]$$

What is introduced in quantum mechanics as a basic postulate will then be a natural theorem of probability theory. Now it follows to advance plausible arguments that the quantum mechanical postulate of Equation 1 has an analogue that appears naturally in probability theory. In Table 1, the analogy between the observables produced by two operators in quantum mechanics and in value theory is portrayed. The position operator, x_i, in quantum mechanics produces the observable leading to the expected measurement of position of an atomic particle described by the wave function of quantum mechanics. In value theory, under appropriate conditions, this operation will lead to the expected inventory of an i^{th} type of commodity. The mathematical operator, $\partial/\partial x_i$, in quantum mechanics leads to a mean value of the measurement of momentum in the i^{th} coordinate; in value theory[9] it leads to an average measurement of the value of a thing of the type i; i.e., the i^{th} commodity.

Normally, one might expect that the mean position of a system (that is, the expected state that will be revealed by a random observation on the state) is given simply by

$$\bar{x}_i(w_0, s_0, t) = \sum_x x_i P(w_0, s_0; x, t), \qquad [10]$$
$$t > s_0.$$

The expected position in the i^{th} coordinate in phase-space can be considered to be the expected inventory of the i^{th} commodity. Equation 10 is appropriate to the situation

[9] See Footnote 8.

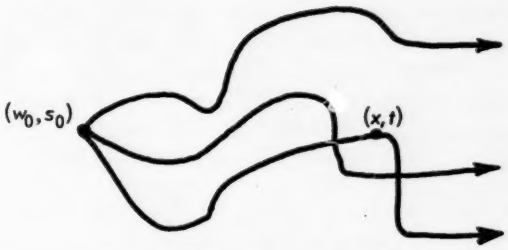


FIG. 2. Illustrating an Average of the First Kind.

Only the initial point (w_0, s_0) is fixed, all possible histories being considered in the average associated with future states of the system.

in which the initial point (w_0, s_0) is fixed *and no further restrictions are considered* (see Figure 2). This is called here an average of the first kind. On the other hand, if both an initial point and a final point are specified, an average of the second kind (see Figure 3) is obtained:

$$\bar{x}_i(w_0, s_0, t, y_0, u_0) = \sum_x P\dagger(x, t; y_0, u_0)\, x_i\, P(w_0, s_0; x, t), \qquad [11]$$
$$u_0 > t > s_0.$$

On the right-hand side of Equation 11 the forward equation gives the probability that the system originating at the point (w_0, s_0) will be in state x at time t. To satisfy the conditions of the average, this is weighted by x_i but it must then be multiplied by the probability that the system moves from state x at time t, to the final fixed point, state y_0, at time u_0. Thus, it is seen that observables associated with operators, as expressed by Equation 1 become a natural consequence in probability theory provided the average is taken at a point intermediate to the two fixed points.

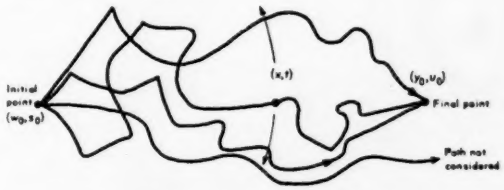


FIG. 3. Illustration of an Average of the Second Kind.

Only those histories that pass through the initial point (w_0, s_0) and the final point (y_0, u_0) are considered. The average is taken with respect to a field point (x, t), which is intermediate; i.e., for which $s_0 \leq t \leq u_0$.

Any physical system described by quantum mechanics fulfills this restriction. The system to which the wave equations apply it is implied in the problem that there is an initiation of the event and an absorption of the event. There are two fixed points in every experiment.

Now consider the space differential operator that in the quantum mechanical system produces an expected measurement of the momentum in the i^{th} direction. The expected value of a thing, $\bar{v}_i$, has been defined as the partial derivative of the value function with respect to the i^{th} commodity:

$$v_i(x, t) = \frac{\Delta Q(x, t)}{\Delta x_i} \qquad [12]$$

If this definition is combined with the fundamental value equation (Equation 11 in Part I), there results Equation 13:

$$\begin{aligned} \bar{v}_i(w_0, s_0, t) &= \sum_x P(w_0, s_0; x, t)\, v_i \\ &= \sum_x P(w_0, s_0; x, t) \frac{\Delta Q(x, t)}{\Delta x_i} \qquad [13] \\ &= -\sum_x Q(x, t) \frac{\Delta}{\Delta x_i} P(w_0, s_0; x, t), \end{aligned}$$

by the nature of the operator adjoint to $\Delta/\Delta x_i$ (see Section VI, D1, Equation 54).[10]

[10] Equation 13 may, of course, be written in terms of the contribution of the k^{th} trapped state (in an absorption system) to the value of a thing:

$$\bar{v}_i(w_0, s_0, t, z_k) = \mu(z_k) \sum_x P(x, t; z_k, \infty) \frac{\Delta}{\Delta x_i} \left\{ P(w_0, s_0; x, t) \right\}$$

obviously,

$$\bar{v}_i(w_0, s_0, t) = \sum_k \bar{v}_i(w_0, s_0, t, z_k). \qquad [13a]$$

In a similar manner the "inventory" average of the second kind may be built up over a set of final states (y_k, u_0):

$$\bar{x}_i(w_0, s_0, t, u_0) = \sum_k \bar{x}_i(w_0, s_0, t, y_k, u_0)$$

If the $\mu(y_k)$ are all unity, then the average of the second kind reduces to that of the first kind. Thus the former summed over only desirable states gives a "useful inventory."

Thus the averages of the type expressed in Equation 1 result naturally in probability value theory. If, indeed, the wave function may be considered a probability function, some of the fundamental relations that appear in quantum mechanics as postulates would result from natural consequences of probability theory.

2. The definition of value and the conservation of momentum. There exists an analogue between the definition of value (Part I, Equation 11) and the law of conservation of momentum. It has been seen above that the negative of the partial derivative with respect to the coordinate as an operator produces analogous quantities in the physical system and the value system—in the first case producing an expected measurement of momentum, and in the second case an expected measure of the value of a thing. For simplicity, assume a simple gambling game. The total money owned by both players is equal to W. The state of the game be represented by the money x owned by the first player. The trapped states, of course, occur when either player wins all the money. Thus the state of the system may be represented as a point on a line segment. It will be assumed that the amount risked in each pot is constant. The value of an added pot to Player 1 is then obtained from the value equation:

$$Q(x) = pQ(x + 1) + qQ(x - 1), \qquad [14]$$

where p represents the probability that Player 1 will win the pot, and q the probability he will lose it. Since $p + q = 1$, the left-hand side of Equation 14 may be multiplied by $p + q$, and rearranged:

$$\begin{aligned} p[Q(x) - Q(x + 1)] \\ = -q[Q(x) - Q(x - 1)]. \end{aligned} \qquad [15]$$

The portion in brackets on the left-hand side represents the difference in state value by an added pot, and therefore by definition, is the value of an increased pot to Player 1. Similarly, the portion in brackets on the right-hand side is the value of a decreased pot to Player 1. This difference taking over a single space interval is, in different terminology, what corresponds to the partial de-

rivative in a continuous space. Thus, Equation 15 is of the form:

$$p \left.\frac{\partial Q}{\partial x}\right|_{+} = -\, q \left.\frac{\partial Q}{\partial x}\right|_{-}. \qquad [16]$$

Since the partial derivative operator is analogous to momentum, Equation 16 may be interpreted under conditions where winning and losing are equal; i.e., where $p = q = \frac{1}{2}$. Here, momentum is conserved; momentum in the positive direction is equal and opposite to momentum in the negative direction. This conservation, of course, is true only in a uniform space. When p is not equal to q, Equation 16 gives the weighting appropriate to a more general expression. Thus it appears that the definition of value, in value theory, has an analogous position to a more general form of the law of conservation of momentum in physical theory.

It should be pointed out that to complete the analogy between probability value theory and physical theory, an important postulate is required in value theory that is missing in this discussion.

3. Analogy with the Special Theory of Relativity. A most challenging task in establishing confidence in the analogical conformity principle is that of demonstrating that the form of physical theory is homomorphic to the form of probability value theory. One's interest is naturally directed to the physical sciences. Since they deal with the simplest of perceptual properties, they are the farthest advanced in the conceptual and analytical sense. As an initial step it will be attempted to show that the simplest of stochastic processes—the random walk—when transformations are made on its transition probabilities in such a manner as to produce uniform motion of the first moment of its probability distribution, will exhibit properties similar to those in the Special Theory of Relativity. Indeed, this transformation is the Lorentz-Einstein transformation referred to the space defined by the stochastic process. Generalization to more complex stochastic processes will be indicated. The demonstration will proceed in three parts. The first will produce evidences of relativistic-like properties of the random walk, the second will introduce a notion of *subjective normalization*, which will lead naturally to the transformation equations. It will be necessary to introduce some concepts associated with time transitions purely on the basis of symmetry and completeness. These concepts will lead to an interpretation of an imaginary component of transition probability. All the basic arguments will be included, although the algebraic manipulations will be minimized to conserve space.

a. Indications of relativistic properties of the random walk. Consider a random walk on one dimension. Let k refer to an index of a space point along a line, and n an index of the n^{th} transition. Only discrete points on a line will be occupied. The probability of finding the system in state k after the n^{th} transition, given that it was known to be in state k_0 at the $n_0{}^{\text{th}}$ transition, is designated by $P(k_0, n_0; k, n)$. The initial space point and transition may be defined as the origin in space and transition indices so that it is understood that $P(k, n)$ represents $P(0, 0; k, n)$. Let the probability of a transition to the next adjacent space point in the positive direction be denoted by p, opposite by q, such that $p + q = 1$, then

$$P(k, n) = pP(k-1, n-1) + qP(k+1, n-1), \qquad [16]$$

for

$$P(0, 0) = 1.$$

This has a solution:

$$P(k, n) = \binom{n}{\frac{n-k}{2}} p^{\frac{1}{2}(n+k)} q^{\frac{1}{2}(n-k)}, \qquad [17]$$

$$n \geq k \geq -n,$$

where $\frac{1}{2}(n + k)$ must be an integer, i.

The first moment $\bar{k}(n)$ is defined by

$$\bar{k}(n) = \sum_k kP(k, n), \qquad [18]$$

and the second moment by

$$\overline{k^2(n)} = \sum_k k^2 P(k, n). \qquad [19]$$

The standard deviation of the distribution, σ^2, is given by:

$$\sigma^2 = \bar{k}^2 - (\bar{k})^2. \qquad [20]$$

There results the average position:

$$\bar{k}(n) = n(p - q), \qquad [21]$$

and the square of the standard deviation

$$\sigma^2(n) = 4npq. \qquad [22]$$

A system that undergoes no change cannot define a space or a time. The actual change itself—the event—is taken to define simultaneously an increment of space and an increment of time. In a system completely isolated from all other systems, time will progress only during the moments of change in that system. Therefore, one is led to the adoption of the index n of the transition number as the measure of the elapsed time. A transition in the system not only defines an ultimate quantum of time, it also defines an ultimate quantum of space. A transition, defining both, can at most move the phase point one unit of space during one unit of time.[11]

The first moment of the distribution gives the "center of probability" of the random walk and therefore its location may be considered the mean position of the stochastic process. Dividing Equation 57 through by the time index n, one gets

$$\frac{\bar{k}(n)}{n} = p - q \equiv \beta. \qquad [23]$$

The ratio of the spatial movement of the center of probability to the time index is equal to the difference between the probabilities of moving to the right, minus the probability of moving to the left. By definitions of time in space, as generated by the system, this ratio, β, becomes the average velocity of the stochastic system. If the probability of moving to the right, p, is greater than that to the left, q, the center of probability will move off to the right with a uniform velocity, β. This velocity will furthermore have a maximum value when $p - q = 1 (q = 0)$. This property is characteristic of the motion of physical systems, where the maximum velocity in physical space is that of the propagation of light *in vacuo*.

[11] "There is no Nature apart from transition, and there is no transition apart from temporal duration" (7, p. 46).

Now consider the first moment and the standard deviation of a random walk that is at rest with respect to the coordinate system. Such a system will be at rest if $p = q = \frac{1}{2}$. Denote the condition at rest by the subscript zero ($_0$), getting:

$$\left.\begin{aligned} \bar{k}_0(n) &= 0, \\ \sigma_0^2(n) &= n. \end{aligned}\right\} \qquad [24]$$

The ratio of the size of the random walk in motion ($k \neq 0$) and that of one at rest ($k = 0$) is given by

$$\frac{\sigma(n)}{\sigma_0(n)} = 2\sqrt{pq}. \qquad [25]$$

We have noted that the velocity is given by $\beta = p - q$; and since $p + q = 1$:

$$\left.\begin{aligned} p &= \tfrac{1}{2}(1 + \beta), \\ q &= \tfrac{1}{2}(1 - \beta), \\ \text{and} \\ \frac{\sigma}{\sigma_0} &= \sqrt{1 - \beta^2}. \end{aligned}\right\} \qquad [26]$$

It is observed that the ratio of the standard deviation of the system in motion to the standard deviation when stationary is equal to the square root of $1 - \beta^2$. Now the standard deviation may be taken as a measure of the physical size of the stochastic process. Hence, one derives that the size of the process contracts in the direction of motion, similar to the Lorentz contraction in the Special Theory of Relativity.

Although the random walk as here described has two of the properties of relativity, this description is not relativistic, since it refers to an absolute frame of reference. Is there a transformation that will express the moving system in terms of its own moving coordinate system, and does it have relativistic properties?

b. Random and deterministic components of probability. The word "random" implies uniform probabilities. In this meaning of the term, random walk implies a uniform probability of movement right or left, i.e., $p = q = \frac{1}{2}$. On the other hand, a deterministic sequence of events has been defined as a sequence in which one state follows the other with a probability of unity. The

stochastic process of the preceding section can be either of these two extremes. If $p = q = \frac{1}{2}$, it is its most random condition; if $p = 1$, $q = 0$, the motion of the phase point is to the right in a fully deterministic fashion. Thus a system whose center of probability is at rest is completely random; and a system whose center of probability moves off with maximum velocity is completely deterministic. This suggests that motion intermediate between these two extremes can be broken down into two components—one that it is random and one deterministic.

We therefore seek to break up the transition probabilities in each dimension into two components, as in Figure 4. The excess of probability of motion to the right over that to the left is defined as the *deterministic component*, which is positive to the right. This leaves a *random component*, q to the right, and $-q$, which is therefore directed to the left; these being equal and opposite, are appropriately termed random.

The deterministic component of probability is a vector quantity. The random component, being added both positively and negatively, sums to zero algebraically. The sum of p and q will be defined as the normalized component and depends directly on the metrics of time and space.

It becomes increasingly clear that the concept of probability is inextricably connected with the constructs of time and space. Obviously, then, any operation affecting the metrics of time and space must also affect the constructs of probability. By the term "event" is meant that something has happened; that is, something has happened with a probability of unity. The probabilities associated with an event are usually normalized to unity. This requirement need not necessarily be adhered to. One can associate with an event an expectation not necessarily of certainty so that the sum of the probabilities is different from unity.

Consider now that time is a dimension in addition to the cartesian coordinates. In the description of the foregoing, it appears that a stationary random walk is described as completely random. On the other hand, it has been assumed that every transition

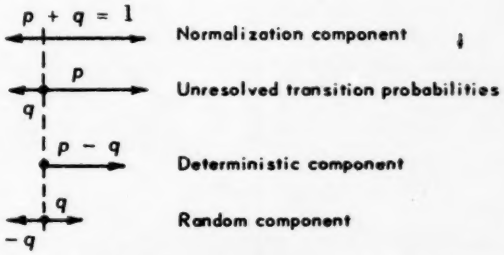


FIG. 4. The Resolution of Transition Probabilities into Deterministic, Random, and Normalization Components.

in space, whether forward or backward, is accompanied in a deterministic fashion (with expectation $p + q$) by transition in time. As long as an absolute frame of reference is adhered to, the transition in space is accompanied by a deterministic transition in time. We now seek to determine what happens to this relation when the coordinate system itself is transformed to move with a velocity equal to the movement of the center of probability of the probability distribution of the stochastic process.

c. Subjective normalization. Consider a universe in which all the denizens themselves, together with what they construe to be materialistic quantities, are all stochastic processes in the same space. Obviously, an observer who is himself a stochastic process, will be subjected to the same distortions to which a local process may be subjected. This is in fact the definition used herein of a *local observer*. He is not only at rest with respect to the center of probability of the process he is observing, but he himself is subjected to the same anisotropies of the probability distributions. Not only is he subjected to the same distortions but he observes the stochastic process as a diffuse probability distribution; he is able to infer the expected movement of the phase point, but he cannot observe discrete transitions to which he could associate a concept of a deterministic event. Hence he is at liberty to adjust the metric of his space and time so that he may reach the normalizations of probability that he desires. In other words, it is useless for him to postulate an absolute discrete framework in which the phase point of his stochastic process is located. Since our observer cannot

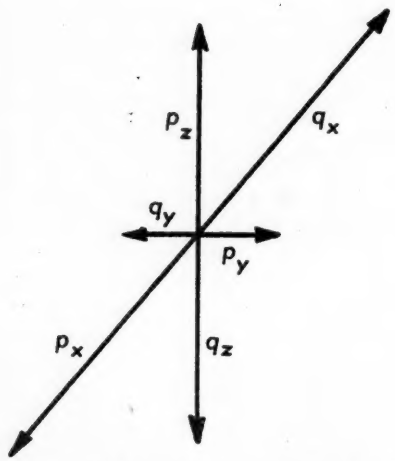


FIG. 5a. Non-isotropic Transition Probabilities in a Stationary Random Walk.

The implications of such a representation made by imposing the observers space metric on the system is that its local time clocks in the three dimensions are asynchronous.

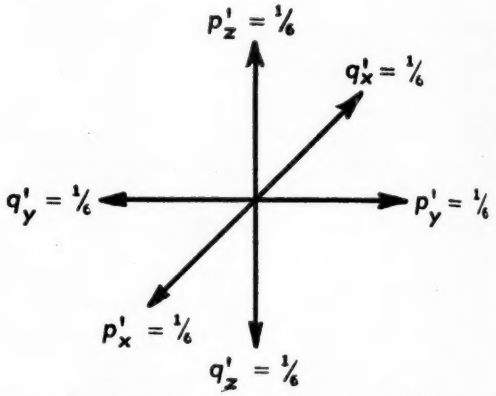


FIG. 5b. The Same Transition Probabilities of Figure 5a as Seen by a Local Observer.

The former observer must conclude that either the systems time clocks are asynchronous and the space metrics of the two systems are identical; or he agrees with the local observer that the time clocks are synchronous (only one time dimension) but then he must conclude that the space metrics of the two systems are different.

conceive of himself as an anisotropic, it would be absurd, he reasons, to believe that his body changes shape merely because it is rotated in space, he sees himself as an isotropic individual. A second observer, presumably ourselves, may see him as flat-headed or pinheaded. We may therefore construe the transition probabilities that define his space to be anisotropic. The distortion is not observable to him, since, to his way of thinking, he is perfectly isotropic. His world may be described in one of three alternative ways: (*a*) we can say that the transition probabilities in the x, y, and z direction are not equal (throughout this part of the discussion we are considering that the process is at rest with respect to ourselves and to the local observer) as in Figure 5a. The local observer, however, considers all of these transition probabilities equal, and equal to $\frac{1}{6}$, as in Fig. 5b.

Now what are the implications of these interpretations? Since time is to be defined in terms of change, the external observer who sees an anisotropic probability distribution *is forced to the conclusion that the clock runs at different rates in different dimensions*, since, for the dimension where the transition probabilities are smaller than the others, transitions will occur less frequently. If the transition in this dimension alone is what determines the movement of its clock, then obviously one is confronted with three dimensions of time—one complementary to each space dimension and all of them occurring at different rates. This interpretation has been brought about by requiring the space metric of the observed system to conform to that of the external observer and leads to the anomalous interpretation of three-dimensional time.

The local observer, on the other hand, subjected as he is to the same anisotropism, cannot observe what we regard to be the distortions of his space. He believes himself to be nicely squareheaded. Not only are his transition probabilities equal in all directions, but the time clocks in all dimensions are synchronized. He can regard time as being a single dimension. There remains a third interpretation, which is the second made by the external observer—ourselves. Instead of imposing our own metric of space on the observed stochastic process and thus arriving at the anomalous description of time, we may wish to preserve the synchronism of the clocks in the three dimensions and infer that the metrics of space of the observed system are different from our own. These metrics of space may be ad-

justed such that the expected motion as observed in the system in the various dimensions agrees with our original interpretation. *This is the meaning of subjective normalization.* An event is interpreted as an expectation. Thus $p + q$, which in the local system is equal to unity, is reinterpreted to mean that $p + q$ multiplied by a space metric is an expectation. It is the expectation that is invariant to these transformations between systems of various isotropism.

In the universe taken as a model for this discussion there is no such thing either as an absolute frame of reference, or an absolutely isotropic system. We have, moreover, inferred there is a separate time dimension concomitant with each space dimension. It is only through the means of subjective normalization that the three time dimensions are synchronized and therefore may be considered as one.

The gist of these remarks is that the normalized component of probabilities in any particular dimension will transform such that the expectation is constant. In other words, that

$$|dx| = (p + q)\,dt = (p' + q')\,dt. \quad [27]$$

If the unprimed probabilities are normalized then one gets an equation of the form

$$dt = (p' + q')\,dt'; \quad \text{or } p' + q' = \partial t/\partial t'. \quad [28]$$

Now consider what happens to the deterministic component of probability. Again assume that the expected motion is invariant to a transformation. The expected motion, dx, is equal to the velocity times dt,

$$dx = \beta\,dt, \quad [29]$$

but the velocity is a deterministic component. Thus, one gets

$$dx = (p - q)\,dt = (p' - q')\,dt'; \text{ or } p' - q' = \partial x/\partial t' \quad [30]$$

for the transformation of a deterministic component as implied by the principle of subjective normalization.

d. The probability transformation. We conconsider two coordinate systems at uniform motion with respect to each other, one designated as the Σ and the other the Σ' system. Consider two random walks, one stationary in the Σ system, one stationary in the Σ' system. Four ways of observing the systems may be considered. The system stationary in the Σ system is observed by an observer local to the Σ system; the same system is observed by local observer in the Σ' system, and vice versa. The following notation will be adopted: (1) The probabilities of transition in Σ as seen by a local observer in Σ, are denoted p_{xx}, q_{xx}. The standard deviation of this system denoted as σ_{xx} and it is measured in terms of the metric of the Σ system. (2) The probabilities of transition in the Σ system as seen by an observer local to the Σ' system, $p_{xx'}$, $q_{xx'}$, $\sigma_{xx'}$. The latter is measured by the local observer in Σ' using the metric of the Σ system. (3) The probabilities of transition in the Σ' system as seen by an observer local to the same system. $p_{x'x'}$, $q_{x'x'}$ and $\sigma_{x'x'}$. The latter refers to the standard deviation seen by the local observer in Σ', and measured by the metrics of Σ'. (4) The probabilities of transition in the Σ' system as seen by an observer local Σ system. $p_{x'x}$, $q_{x'x}$, $\sigma_{x'x}$, the latter being the standard deviation observed by the local observer in Σ', measured using the metrics of the Σ' system. Now any measurement of space using the same metrics of space of the same phenomena should result in the same numerical result. Thus we conclude

$$\left.\begin{aligned} \sigma_{xx'} &= \sigma_{xx} \\ \sigma_{x'x'} &= \sigma_{x'x} \end{aligned}\right\}, \quad [31]$$

and therefore consider as a fundamental invariant the preservation of the standard deviation, or rather the square of this deviation, when measured by the same space metric:

$$\left.\begin{aligned} 4\,p_{xx}q_{xx} &= 4\,p_{xx'}q_{xx'} \\ 4\,p_{x'x}q_{x'x} &= 4\,p_{x'x'}q_{xx'} \end{aligned}\right\}. \quad [32]$$

Consider now a stochastic process stationary in the Σ system. The transformation to the Σ' system should result in its uniform motion with respect to the Σ' system, and if the conditions of relativity prevail it will appear to be foreshortened in the Σ' system. Conversely by the fundamental principle of relativity the system that is stationary

in the Σ' system when transformed to the Σ system, will appear again to be foreshortened. The original conditions of the systems stationary in the Σ' system—call it process *I*—is

$$\left.\begin{aligned} p_{xx} + q_{xx} &= 1, \\ p_{xx} - q_{xx} &= 0. \end{aligned}\right\} \qquad [33]$$

etc. for y and z.

The transformation to the Σ' system will be undertaken in two parts. In the first step the external observer sees the other stochastic system as if it were non-uniform and the probabilities not normalized in an isotropic space. By inference this non-uniformity implies the observed time dimensions are asynchronous. The second step distorts the space in a manner which restores normalization and synchronizes the clocks in the three dimensions; thus permitting time again to be regarded as a single dimension. In the first step, Equation 33 will be transformed in the form that an observer local to the Σ' system would view the same probabilities if he were employing the metrics of space and time appropriate to Σ system. Applying the rules for subjective normalization previously expounded there results, letting r, s, t, be three time coordinates corresponding respectively to x, y, z:

$$\left.\begin{aligned} p_{xx'} + q_{xx'} &= \partial r/\partial r' = e_{rr'} \\ p_{xx'} - q_{xx'} &= \partial x/\partial r' = e_{xr'} \end{aligned}\right\} \qquad [34a]$$

$$\left.\begin{aligned} p_{yy'} + q_{yy'} &= \partial s/\partial s' = e_{ss'} \\ p_{yy'} - q_{yy'} &= \partial y/\partial s' = e_{ys'} \end{aligned}\right\} \qquad [34b]$$

$$\left.\begin{aligned} p_{zz'} + q_{zz'} &= \partial t/\partial t' = e_{tt'} \\ p_{zz'} - q_{zz'} &= \partial z/\partial t' = e_{zt'} \end{aligned}\right\} \qquad [34c]$$

Equations 34a, b, c, together with the invariances of Equations 32 and 31, are almost a complete set of conditions for an orthogonal transformation in the space-time framework, i.e.

$$\left.\begin{aligned} e^2_{rr'} - e^2_{xr'} &= 1, \\ e^2_{ss'} - e^2_{ys'} &= 1, \\ e^2_{tt'} - e^2_{zt'} &= 1. \end{aligned}\right\} \qquad [35a]$$

The missing conditions are

$$\left.\begin{aligned} e^2_{xx'} - e^2_{rx'} &= 1, \\ e^2_{yy'} - e^2_{sy'} &= 1, \\ e^2_{zz'} - e^2_{iz'} &= 1, \\ e^2_{\alpha\alpha'} &= e^2_{\beta\beta'}, \\ e^2_{\alpha\beta'} &= e^2_{\beta\alpha'}, \end{aligned}\right\} \qquad [35b]$$

where α and β are commuted through the six dimensions.

We shall assume *ad hoc* that Equations 35b are true[12] and attempt to reconstruct a situation which will produce them. This reconstruction is not unique without the addition of conditions imposed by some deeper insight—which is missing here. Such an insight may be forthcoming if a successful extension of this analogy to the General Theory of Relativity is accomplished. Equations 35b imply that there is a set of probabilities associated with transitions in time, p_{rr}, q_{rr}, etc., for which

$$4\, p_{rr'}q_{rr'} = (\partial x/\partial x')^2 - (\partial r/\partial x')^2, \text{ etc.}$$

is an invariant and equal to unity.

Now $p_{xx'} + q_{xx'}$ is the expectation that any space (either positive or negative) transition occurs per *deterministic time transition*. The difference $p_{xx'} - q_{xx'}$ is the expectation that a *deterministic* space transition occurs per *deterministic* time transition. Hence $(p_{xx'} - q_{xx'})/(p_{xx'} + q_{xx'})$ is the velocity of the system. The interpretation of $p_{rr'} - q_{rr'}$ cannot be specifically made on the basis of Equations 35b alone. It seems natural to consider it a deterministic change in time and hence should be equal to unity for a stationary system. Some other interpretation, such as a fully symmetrical one, may well be the better one however. We were led to try the following, since the imaginary notation fits condition 35b and is consistent with the requirement of rotation of space coordinates into an imaginary time dimension; this latter is the result of the invariance of the difference between components in Equations 35b:

$$\left.\begin{aligned} p_{rr'} + q_{rr'} &= i\partial r/\partial x' = ie_{rx'} \\ p_{rr'} - q_{rr'} &= i\partial x/\partial x' = ie_{xx'} \end{aligned}\right\} \qquad [36a]$$

[12] The space-time probabilities producing Equations 35a are insufficient to define a relativistic space-time metric. The adoption of Equations 35b is then the generalization on probability which produces this characteristic.

$$\left.\begin{aligned} p_{ss'} + q_{ss'} &= i\partial s/\partial y' = ie_{sy'} \\ p_{ss'} - q_{ss'} &= i\partial y/\partial y' = ie_{yy'} \end{aligned}\right\} \quad [36b]$$

$$\left.\begin{aligned} p_{tt'} + q_{tt'} &= i\partial t/\partial z' = ie_{tz'} \\ p_{tt'} - q_{tt'} &= i\partial z/\partial z' = ie_{zz'} \end{aligned}\right\} \quad [36c]$$

This results in a six-dimensional transformation.

Conditions 35a, b, now imply an orthogonal transformation which can be characterized as a rotation of space axes into the corresponding imaginary time axes. Aligning the x-axis with the direction of motion simplifies this somewhat. The transformation will involve only one dimension of space and of time. Anticipating that the second part of the transformation will restore the synchronism of the time-dimension clocks, Equations 34 and 36 may be rewritten in a four-dimensional space, assuming only one time dimension. There results, letting $r = s = t$:

$$\left.\begin{aligned} p_{xx'} + q_{xx'} &= \partial t/\partial t' = e_{tt} \\ p_{xx'} - q_{xx'} &= \partial x/\partial t' = e_{xt'} \end{aligned}\right\}, \quad [37a]$$

$$\left.\begin{aligned} p_{yy'} + q_{yy'} &= 1 \\ p_{yy'} - q_{yy'} &= 0 \end{aligned}\right\}, \quad [37b]$$

$$\left.\begin{aligned} p_{zz'} + q_{zz'} &= 1 \\ p_{zz'} - q_{zz'} &= 0 \end{aligned}\right\}, \quad [37c]$$

$$\left.\begin{aligned} p_{tt'} + q_{tt} &= i\ \partial t/\partial x' = ie_{tz'} \\ p_{tt'} - q_{tt'} &= i\ \partial x/\partial x' = ie_{xz'} \end{aligned}\right\}. \quad [37d]$$

Several considerations contribute to the addition of Equations 34 and 36. Before completing this part of the explanation, consider what happens when the second state of the transformation is conducted. The second state of the transformation results when the observer local to the Σ' system imposes normalization, and one gets:

$$\begin{aligned} p_{x'x'} + q_{x'x'} &= \frac{p_{xx'} + q_{xx'}}{e_{tt'}} = 1 \\ p_{x'x'} - q_{x'x'} &= \frac{p_{xx'} - q_{xx'}}{e_{tt'}} = \frac{\partial x}{\partial t} = \beta_x \end{aligned} \quad [38]$$

The subjective normalization simultaneously normalizes the probabilities in the appropriate dimension—leaves the deterministic component that of a velocity—and synchronizes the clocks in the various dimensions reducing an apparently six-dimensional space to four dimensions. Now consider what effect these stages of the transformation have upon the standard deviation. The first stage of the transformation, according to Equations 32 and 35, leaves the standard deviation unchanged. Now the imposition of the second stage of the transformation which resulted in the normalization of the probability components results:

$$\begin{aligned} \frac{4p_{xx'}\,p_{xx'}}{e_{tt'}} &= 4p_{x'x'}\,q_{xx'}\,, \\ &= (1 - \beta_x)^2 \end{aligned} \quad [39]$$

thus

$$\frac{\sigma_{x'x'}}{\sigma_{xx}} = \sqrt{1 - \beta_x^2}, \quad [40]$$

i.e., the Lorentz foreshortening as previously encountered. These conditions are now those of an orthogonal transformation in the four-dimensional system from the Σ to the Σ' frames of reference. They may be interpreted as a rotation of the axis of motion, say for example, the x axis, into an imaginary time axis. The imaginary is associated with the probabilities relating to the time transition in Equations 34 and 36 in order to preserve the same form of equation in Equation 35. It results, in effect, in a more consistent notation. The $e_{xx'}$ etc. may also be interpreted as a projection of unit vectors in the Σ system on the Σ' axes. The Lorentz-Einstein transformation is implied by the process described. Since the difference of the squares is invariant rather than their sum, the rotation of coordinates must be through an imaginary angle. Assume the rotation to be of the x-axis into the imaginary time axis (uniform motion along the x-axis), this imaginary rotation and the imaginary notation make it appropriate to consider that the unit vector in the time dimension is also an imaginary. The sum of the time transition probabilities $p_{tt'} + q_{tt'}$ has been set equal to $ie_{tx'}$ and the difference to $ie_{tt'}$ in order that the time transition for a stationary system will be completely deterministic. Its projection upon any space axis will be an imaginary. Writing down, now, the conditions for the transformation of the unit vectors in the Σ sys-

tem to the unit vectors in the Σ' system, one gets

$$\begin{aligned} \boldsymbol{e}'_x &= e_{xx'}\boldsymbol{e}_x + (ie_{tx'})(i\boldsymbol{e}_t), \\ \boldsymbol{e}'_y &= \boldsymbol{e}_y, \\ \boldsymbol{e}'_z &= \boldsymbol{e}_z, \\ i\boldsymbol{e}'_t &= (-ie_{xt'})\boldsymbol{e}_x + e_{tt'}(i\boldsymbol{e}_t). \end{aligned} \tag{41}$$

e. The Lorentz-Einstein Transformation Equations. The transformation equations for the unit vector may now be used to express the transformation for any vector quantity. Suppose the four-vector displacement vector $\boldsymbol{R} = \boldsymbol{R}'$ is considered. This four-vector being an expected motion is invariant to four-dimensional rotation transformation, one has:

$$\begin{aligned} \boldsymbol{R} &= \sum_j x_j \boldsymbol{e}_j \\ \boldsymbol{R}' &= \sum_i x_i \boldsymbol{e}_i \end{aligned} \tag{42}$$

Where $\boldsymbol{e}_x$, etc. represent the unit vectors and x_j with $j = 1, 2, 3, 4$ replaces x, y, z and it.

One has, now, substituting the transformation for the unit vector:

$$\boldsymbol{e}'_i = \sum_j e_{ji}\boldsymbol{e}_j \tag{43}$$

getting

$$x_j = \sum_i x'_i e_{ji} \tag{44}$$

Substituting now into Equation 44 the appropriate coefficient selected from Equation 41, there results

$$\begin{aligned} x &= e_{xx'}x' - e_{xt'}t', \qquad y = y' \\ t &= e_{tx'}x' + e_{tt'}t' \qquad z = z' \end{aligned} \tag{45}$$

or, making note of the invariancies of Equation 40, and the relations outlined in Equations 38 and 39, one gets the well-known Lorentz-Einstein transformation equations of the special theory of relativity as referred to the space of the stochastic process:

$$\begin{aligned} x &= \frac{x'}{\sqrt{1-\beta^2}} - \frac{\beta t'}{\sqrt{1-\beta^2}} \\ t &= \frac{-\beta x'}{\sqrt{1-\beta^2}} + \frac{t'}{\sqrt{1-\beta^2}}, \end{aligned} \tag{46}$$

Equations 45 and 46 are then a more general statement of the principle of subjective normalization elucidated in Equations 27, 28, 29, 30:

$$\begin{aligned} dt &= (p_{xx'} + y_{xx'})dt' + i(p_{tt'} + q_{tt'})dx' \\ dx &= (p_{xx'} - q_{xx'})dt' - i(p_{tt'} - q_{tt'})dx' \end{aligned}$$

We have had to resort to arguments of consistency and symmetry to bring in the probabilities associated with the transitions in the time dimension. Since the theory of relativity itself is so firmly established and since an actual stochastic process in real space would be subject to the conditions giving rise to this special theory of relativity, one may regard that theory itself as supporting some of the results here obtained. The equation for the relativistic retardation of the time will be produced by applying the condition of subjective normalization to the second half of Equation 40. It will also result, of course, from the resultant Lorentz-Einstein transformation equation.

The imaginary in Equations 34 and 36 was introduced in order to preserve the form of Equation 40 and to represent $p_{tt} - q_{tt'}$ as a determinate time transition. This provides a possible explanation for the interpretation of an imaginary transition probability. In the four-dimensional complex the complete description of the stochastic motion requires not three sets of branching probabilities, but four, the fourth one being imaginary. One set of three is associated with space-time transitions, and another set of three which has been reduced to a single entity through subjective normalization and which are associated with the time-space transitions.

f. Non-random walk system. The question naturally rises now, of how to generalize these remarks to apply to stochastic processes not of the simple random walk character, particularly to stochastic processes, the transition probabilities of which vary from point to point in space. Consider three systems, Σ, Σ', and Σ'', moving respectively with velocities zero, β_{01}, i.e., the motion of Σ' with respect to Σ), β_{02} i.e., the motion of Σ'' with respect to Σ) and β_{12} (i.e. the motion of Σ'' system with respect to Σ'). These

correspond to rotation of the coordinates by the angle ψ_{01}, ψ_{02} and ψ_{12}, where

$$\psi_{02}, = \psi_{01} + \psi_{12} \qquad [47]$$

or

$$\left.\begin{aligned} \tan \psi_{02} &= i\frac{e_{xt'}}{e_{xx'}} = i\beta_{02}, \\ \tan \psi_{01} &= i\beta_{01}, \\ \tan \psi_{12} &= i\beta_{02}, \end{aligned}\right\} \qquad [48]$$

and since

$$\tan \psi_{02} = \frac{\tan \psi_{01} + \tan \psi_{02}}{1 - \tan \psi_{01} \tan \psi_{12}}, \qquad [49]$$

one obtains

$$\frac{p_{xx''} - q_{xx''}}{p_{xx''} + q_{xx''}} = \frac{\dfrac{p_{xx'} - q_{xx'}}{p_{xx'} + q_{xx'}} + \dfrac{p_{x'x''} - q_{x'x''}}{p_{x'x''} + q_{x'x''}}}{1 + \dfrac{p_{xx'} - q_{xx'}}{p_{xx'} + q_{xx'}} \dfrac{p_{x'x''} - q_{x'x''}}{p_{x'x''} + q_{x'x''}}}. \qquad [50]$$

Equation 50 may be used to transform the transition probabilities point by point in such a manner that relativistic invariancy is preserved.

C. The Uncertainty Principle

An important result appearing in quantum mechanics is the Heisenberg Uncertainty Principle (7).[34] The Heisenberg Uncertainty Principle states that the ultimate residual errors associated with the simultaneous measurement of two canonically conjugate variables is greater or equal to $h/4\pi$ where h is Planck's constant. More exactly, the uncertainty principle is stated as follows:

$$(\overline{\Delta p})^2 \cdot (\overline{\Delta q})^2 \geqq -\frac{1}{4}\left\{\int \phi^*(O\bar{O} - O\bar{O})\phi \, d\tau\right\}, \qquad [51]$$

where p and q are canonically conjugate variables, O is the operator corresponding to p and $\bar{O}$ the operator corresponding to q. ϕ is the state function and ϕ^* its adjoint —or in the case of quantum mechanics, its complex conjugate. Now, the simple operators pertaining to canonically conjugate variables bear a simple relationship to each other: if x_i is a coordinate operator, which on operating on the scalar product produces an expected measure of position, $\partial/\partial x_i$ is the operator that produces a quantity that is canonically conjugate to the first one. In physics, the implication of the uncertainty principle, for example, is that the position and momentum of a particle may not be measured simultaneously with infinite accuracy; as the position is pinned down more and more exactly the uncertainty in the estimate of the momentum increases without bound, and vice versa. Strictly speaking the uncertainty principle in quantum mechanics, which incidentally also appears in sound theory, is applied to wave motion. In its most general aspects this restriction is probably not as necessary as, for example, the appropriate generalization of the reference above. In the terminology of value theory the operator x_i produces an expected measurement of an inventory of a system that is intermediate between two fixed conditions, and the operator $x_i = \partial/\partial x_i$ produces an expected measurement of the value of one thing of type i. Thus *the uncertainty principle in value theory would state that on an empirical basis one cannot simultaneously take an inventory of things of type i and know the value of these things with infinite accuracy.* Empirical values are determined in the market place.

For example, suppose we are among the early Dutch settlers at their first arrival at the island of Manhattan. We have in our possession several trunks full of colored beads that we intend to trade to the savages for their land. How much land will a hatful of the beads buy? In other words what is the value of the beads in this context? We know exactly how many beads we have, but know nothing about their value; so an expedition sets out, trades off a small portion of the beads for land and thus an estimate of their value is obtained. Two things have occurred: one, the market value for beads has been affected; secondly, we have been forced to consume part of the beads and hence did not have as many as we had before. The only way one can deter-

mine the value of all the beads is by trading them all off; but then, of course, there aren't any more. As a second example we ask what is the value of a cake that is placed before us? Obviously a sample bite must be taken to estimate its value. But in the process of eating part of the cake our appetite is affected and the last bite will not be of the same value as the first. The only way in which the value of the entire cake may be measured is by eating it all. But we can't determine the value of the cake by eating it and also have it. This is the uncertainty principle. Applying operators pertaining to value theory but similar to Equation 51 produces:

$$\overline{(\Delta v_i)^2} \cdot \overline{(\Delta x_i)^2} \geq \frac{1}{4}\left[\sum_x P^+\left(\frac{\partial}{\partial x_i}x_i - x_i\frac{\partial}{\partial x_i}\right)P\right]^2 = \tfrac{1}{4}\left[\sum_x P^+P\right]^2 = \tfrac{1}{4}Q(x_0). \quad [52]$$

Again, these arguments are to be considered plausibilities since the rigorous investigation has not been completed.

D. Analogical conformity in other disciplines

The purpose of this section is to call attention to other disciplines where the principle of analogical conformity may be examined and ultimately exploited.

1. Adjointness in economic models. It has been stated that the adjoint function may always be interpreted as a value function. In most systems the forward function is said to be "real" in that there may be an observable flux associated with it. In the general case of value theory the history of systems is observable and, with a sufficiently large number of observations, the time dependency of state probability may be inferred. In the case of the diffusion of molecules or neutrons, for example, the diffusion may be observed by appropriate detection devices. In such systems the adjoint function has the character of a mathematical construct. In an economic system, however, the situation is different. There is not only a flux of real things diffusing throughout the states of the system, but there is also an adjoint commodity that diffuses in a manner described by the adjoint equation. This commodity is money.

An operator O, which belongs to a class of functions η_i, is said to be adjoint to an operator $O\dagger$, which belongs to a class of functions ζ_j, if:

$$(\zeta_j \cdot O\eta_i) = (O\dagger\zeta_j \cdot \eta_i) \quad [53]$$

In eigenvalue problems this restriction ensures that the eigenfunctions of O are bi-orthogonal to the eigenfunctions of O^+, a condition that is useful in expanding an arbitrary function into terms of the eigenfunctions. As a result of this condition, a multiplicative operator x, is self-adjoint, that is, $x = x\dagger$. One can show by application of Equation 53 that where ξ is any coordinate including time, the operator partial derivative with respect to ξ is replaced by its negative in order to produce the appropriate adjoint operator:

$$\partial/\partial\xi = -(\partial/\partial\xi)\dagger \quad [54]$$

Second derivatives are self-adjoint, and it may be shown that a source in the forward system is replaced by a sink in the adjoint system, and vice versa. The replacement of the first derivative by the negative indicates that velocities and other rates of change in the adjoint system are reversed. Hence the adjoint system is often called the backward system. On the other hand, since the second derivatives are self-adjoint, diffusion that takes place in the forward system will take place similarly in the adjoint system. The movement of money obeys all these requirements. Money is always exchanged for commodities, hence its motion is in the opposite direction. It can be expected to diffuse similarly to commodities. And lastly, a source of money is a sink of commodities, and vice versa. This suggests that first-order perturbation theory can be exploited in economic theory.

2. The second law of thermodynamics. The second law of thermodynamics—so important in physical science—can be expected to describe the behavior of the statistical and stochastic formulation of trial coalitions

in many-person game theory. The coalitions so produced gradually move toward an equilibrium configuration subject to the rules of interaction. This equilibrium cannot be reached instantaneously, since every participant of the game does not interact simultaneously with all other participants. Chance meetings occur and trial coalitions are considered. If it is to the mutual advantage of the players involved in the meeting, new coalitions may develop. These trials go on continuously, more often resulting in no change in the coalition structure. The availability of unusual values resulting from shifts of coalitions will gradually disappear or become a minimum as equilibrium is reached. A function—similar to entropy in logical form—i.e., a logarithm of a probability function—will continually increase toward a maximum at equilibrium. The principle of analogical conformity can be expected to give guidance to the application of old well-established formalistic principles to newly formed theories.

3. Psychiatry. It appears that the description of the evolution of value systems may have some bearing on a description of the technique of psychiatry. One of the first symptoms of psychopathological disturbance is the chronic inability to reach decision; i.e., to conflicts in value structures. The psychiatrist may be said to help his patient in performing his own concrescence, particularly when there lies embedded deep in the patient's value structure underlying schisms that make so difficult his performing his own concrescence. In this terminology it is more appropriate to term these underlying situations as schisms, rather than as conflicts, since the past conflict has been removed by the schism existing. The underlying schism is responsible for the conflict at the present conscious level. It has long been recognized that the healing art of psychiatry cannot be described in terms of a few simple operations. A psychiatrist, in leading his patient to perform a concrescence, is in fact aiding in a highly creative process and one not to be described in terms of a series of operations. It might be suggested that the terminology of psychiatry would be greatly improved in a semantic connotation if the technical language of value theory were adopted to replace a technical language built around the extreme manifestations of primitive urges.

Consider an example; one can recognize three stages of fear. In the first stage there is a rational response to the stimulus giving rise to fear. Not only is there physiological response, but the individual himself makes decisions tending to reduce the danger in the situation faced. The second stage of fear is accompanied by physiological responses to the stimuli, but the power of decision is diminished—or even destroyed completely. A third stage of fear may be reached in which both the physiological responses and decision responses occur but are apparently not correlated necessarily with a situation of real danger to the individual—certain secondary stimuli to which he may have become conditioned may produce the anxiety reaction or he may simply be afraid.

These reactions may possibly have an explanation in the terminology of value theory. Assume a particular individual to be strongly self-oriented, in fact, his ethical postulates with respect to his own survival is such that he has, in effect, postulated his own immortality.

If a stress occurs which threatens his survival, he takes rational action—he makes decisions leading toward a relief of the stress. Suppose, however, that the stimulus is so strong that the individual for the first time becomes aware of his mortality. We have seen that whenever the inevitable death of a system emerges from the unpredictable future that a value system based solely on survival must necessarily collapse to zero everywhere. An individual so oriented, and so stimulated is suddenly confronted with a complete inability to form values, and hence to make decisions. He may react due to his physiological response but not in a rational manner. His rational framework has become disrupted.

What happens in such a situation? The individual usually reforms his value system around some postulates which recognizes the mortal nature of his system. This may occur almost immediately, or may take a few days.

If the collapse of the value structure has been catastrophic the individual is left with no underlying values which he may use as a basis for the reformulation of "mortal" values. He needs a fresh imprintation, and selects his fear experiences for this new imprintation; i.e., all the stimuli he was receiving at the time of his fear experience becomes the basis in an uncorrelated fashion for his new imprintation. Hence he appears "conditioned" in all manner of apparently irrational ways. Any of the stimuli which occurred during his ordeal of fear relevant or irrelevant to the true source of danger will reproduce the conditions motivated by fear. A round ball may frighten him—the sound of friendly aircraft may induce him to dive under cover even though he is many miles away from the combat area, etc. Some of the irrelevant stimuli may be the result of everyday occurrences, everyday sounds, etc. Since these are now imprinted into a new value structure associated with fear, and since these stimuli constantly reoccur, the individual is kept in a constant state of anxiety by these everyday, irrelevant stimuli. Yet he is behaving (i.e., making decisions) with complete consistency with respect to his new set of values.

VII. SUMMARY

It should be apparent that a great deal of work remains to be done in the realm of value theory. In the first place, its connections with other aspects of the theory of decision processes need to be firmly connected. A connection between value theory and learning theory has been entirely absent, although at one point there has been an implied relationship. This is to be found in the relationship between rigidity of value systems and confidence in the workability of value postulates based upon the number of successful times they have been employed for decision purposes. Also absent is a connection between value theory and information and communication theory, although at times the discussion has bordered on this area.

One can anticipate how such a connection will be established. Today's decisions are based upon information at hand, that is, on information that was generated prior to the decision. There are time delays in the communication of this information from its source to the decision-maker. The decision is made and then implementation starts, but implementation is not achieved instantaneously; orders and directives and interpretation of the decision must filter down to all echelons. In both processes there is degradation of information.

The situation is analogous to the solutions of the scalar Maxwell equation in electromagnetic theory. Two solutions result: one called an advance potential and one a retarded potential. The retarded potential has an obvious interpretation. It is a wave generated by the disturbance, which proceeds away from it into space with the velocity of light. The advance potential is sometimes discarded as simply a peculiarity introduced by the mathematical formulation and of no interest. In the light of the foregoing discussion an interesting interpretation can be made of the advanced potential. It is simply the adjoint solution. It represents information arriving at the point of the disturbance at the time of the disturbance, which must have been generated in advance of the action. It is a function by means of which one predicts the importance of the particular action; i.e., the ultimate effect the action has on the external universe. Thus there is always this unavoidable delay between the gathering of information relative to importance and the implementation of the action. In the meantime the universe may be rapidly changing. If so, degradation of rational action will result.

By way of summary it may be stated that a simple stochastic model is assumed as applicable to the world of experience. Decision criteria direct attention immediately to values associated with the system by means of a fundamental value equation; values are fundamentally associated with states of the system and are continually referred to values of future states until some ultimate endpoint is reached or until unpredictability degrades the system into complete randomness. Values for these ultimate states must be postulated in order to arrive at solutions for values of intermediate states. Such values

are intuitive in character. The degradation of retrodiction has been pointed out by von Neumann and by Watanabe. Postulated herein is the degradation of prediction. Rationality seems to pertain to a small island of states adjacent in time and space to the decision-maker and ending diffusely in uncertainty and intuition.

Values of things are given precise definition in the given context. This definition leads naturally to a law of diminishing and increasing returns—to a methodology for marginal analysis and the place of cost and effectiveness in methodologies. The relationship between intrinsic and extrinsic values is defined. The sources of disagreements and uncertainties in the postulation of ultimate values have been investigated, leading to a description of conflict and decision in the means of their resolution. These methods, dominance-suppression, schism, and concrescence, have led naturally into a discussion of the dynamics of the growth of value systems. Such an evolution has required two basic assumptions: that conflict in decision is a fundamental stress producing a tension that must be removed and that simplicity of the fundamental value structure is a desirable quality. In opposition to the tension of conflict there is the rigidity of confidence. Confidence in a value system results from its successful use on a statistical basis.

The notion of confidence is defined as a measure of reality. Next, the formalistic identity between growth of value systems and growth of scientific theories is established, leading to the conclusion that there is no logical difference in the nature of materialistic and moral constructs.

Next, by way of illustration, it is indicated how several ethical systems may be rephrased in the terminology of value theory. One of these in particular, casuistry, is the basis for an experimental approach to ethics. A particular experimental approach raises the question of the absoluteness versus relativeness of the fundamental value constructs. This problem is shown to be of the same nature as the epistemological problem of reference to the reality of the external world, and it results in the same conclusion: an external observer cannot detect the difference in the actions of individuals adhering to one or the other of the viewpoints.

Lastly, the principle of analogical conformity is enunciated and evidences and illustrations leading to its plausibility are set forth. These evidences occur in quantum-mechanical theory, in the special theory of relativity, in the uncertainty principle, in economic models, in psychology, in psychiatry, as well as in the description of the rational process. It is believed that value theory provides a technical language and formalistic approach that may add materially to the unification of the rational disciplines.

REFERENCES

1. Churchman, C. W. *Theory of experimental inferences*. New York: Macmillan, 1948.
2. Churchman, C. W., & Ackoff, R. L. An experimental definition of personality. *Phil. Sci.*, 1947, 14, 304–332.
3. Churchman, C. W., & Ackoff, R. L., with Comments by N. M. Smith, Jr. An approximate measure of value. *J. Oper. Res. Soc. Amer.*, 1954, 2, 172–187.
4. Feller, W. *An introduction to probability and its application*. New York: Wiley, 1950.
5. Leys, W. A. R. *Ethics for policy decisions, the art of asking deliberative questions*. New York: Prentice-Hall, 1952.
6. Lindsay, R. B. & Margenau, H. *Foundations of physics*. New York: Wiley, 1936.
7. Whitehead, A. N. *Nature and life*. Chicago: University of Chicago Press, 1934.

(Manuscript received March 31, 1956.)

I dare say that I have worked off my fundamental formula on you that the chief end of man is to frame general propositions and that no general proposition is worth a damn.

—Oliver Wendell Holmes, Jr.

MIND-BODY DUALISM AND THE UNITY OF BODYMIND

By John Money[1]

The Johns Hopkins University School of Medicine

Like the astronomer who studies "messages" emanating from heavenly bodies or the geologist who studies "messages" recorded in the rocks, the investigator of mind also is concerned with messages, namely those which indicate how possessors of minds decipher the messages they receive. The function of mind becomes, then, for the scientist a deciphering activity of certain classes of living beings. The dilemma of mind-body "dualism" seems to disappear in this context.

TRY one day the experimental game of asking at random a half dozen psychologists and psychiatrists what are the primary data of their work—the bricks with which they figuratively build their houses. The diversity of their answers is surprising and somewhat disconcerting. Cognition, affect, conation; motives, drives, needs; emotions, attitudes, behavior; interpersonal relations, conflicts, adjustments and ego defences—all these and more come in for mention. It is no longer reputable, however, to mention ideas, thoughts, and functioning of the mind.

Perhaps it is unfairly ironical to say that it is disreputable to mention the mind these days. It is more accurate to say that contemporary psychological theory is caught on the horns of a treacherous dilemma with regard to mind and body. On the one hand, psychology is the study of mind. On the other hand, psychology is a biological discipline, and for biologists mind is too vague and intangible a concept to define with scientific vigor. Divided in their allegiance, investigators of the mind are prone to cast around for concepts and terms that at least do not patently flout biology. Psychobiology is one such term; psychosomatic is another. Yet, these double-barrelled terms do not resolve the mind-body dilemma. On the contrary, they tend only to accentuate it, by pointing up the duality. Intended to unite the refractory pair, these terms often succeed only in divorcing them. Whenever men use an abstract noun, that noun soon becomes father of a thing: and so there is a thingness, a misplaced concreteness, about mind as well as body, about psyche as well as soma. Whenever men speak of two things, then they will surely, sooner or later, allow one to be cause, the other effect; and so arises the controversy of whether somatic happenings are caused by the psyche, or vice versa, or both.

A brief historical glance will help to throw some light on the mind-body dilemma. In the seventeenth century, Descartes made his famous distinction between mind and body, postulating the pineal gland as their organ of conjunction. In the theological climate of his time, there was great pragmatic usefulness in the Cartesian distinction. It left the mind and the soul as the province of metaphysics, theology, and supernatural law, while releasing the body and matter as the province of science and the laws of nature. The natural sciences were thus released from ecclesiatical charges of heresy, and scientists were mostly spared the tortures of the inquisition.

Growth and development of the natural sciences proceeded by leaps and bounds in the eighteenth century. Some of the riddles and imponderables of the universe that had haunted mankind were resolved with spectacular success. More and more, it seemed

[1] Assistant Professor of Medical Psychology, appointed in the Department of Psychiatry to do full-time research, supported by a grant from the Josiah Macy, Jr. Foundation, in the pediatric endocrine clinic, The Johns Hopkins Hospital.

certain that the methods and theories of astronomy and physics were the key that would unlock the riddles of biology. Working on this assumption, Darwin in the nineteenth century showed, in his theory of evolution, how to put the key into the lock.

In the post-Darwinian decades of the nineteenth century, it was an easy conjecture that the methods and theories of the material sciences would succeed, where philosophy had failed, in unlocking the secrets of the mind. For the first twenty years of his scientific career, Freud adhered to this conjecture and made a monumental effort to formulate his theories of psychopathology in terms of neurophysiology. But his psychoneuronal theory would not work and was not published. It had become almost more occult and speculative than a theory, formulated without neurophysiological concepts, that used mental or psychic concepts exclusively. By 1900, Freud had abandoned neural mechanisms and, influenced chiefly by the psychology of Herbart and by Kant's famous trilogy of cognition, affect, and conation, he published a nonmaterialist theory of mind in the last chapter of *The Interpretation of Dreams*. The twentieth century opened, therefore, with a full turn of the Cartesian wheel completed. Mind and body were as separate and distinct as they ever had been. There was not even a pineal conjunction for them.

Today, in the middle of the twentieth century, investigation of mind in science and medicine is still bedeviled by an intractable mind-body dualism. One may inquire whether there is any way of circumventing this contemporary predicament. The remainder of this paper is directed toward finding an answer to this inquiry.

The investigator of mind, like every other scientific investigator, begins all of his endeavors by observing signs and signals. He is especially interested in the signs and signals which human beings make or display. Yet, interestingly enough, all scientific investigations of any kind whatsoever are investigations of human signs and signals. To illustrate, one says that an astronomer is interested in signals from the stars, but these signals convey messages which are unconditionally private and uninvestigable until and unless the individual astronomer reports them in some way or another—that is, unless he relays or reports them to other human beings. The astronomer and his colleagues refer the star-message to a far distant source, namely the star, but the message itself is something that happened to the astronomer. If he did not relay it, it would have been so inaccessibly private, that he may as well not have received it, so far as science is concerned.

When a person gets messages, he is entitled to do what he likes with them. If he is an astronomer, he is quite entitled to say: "This visual message was signaled from a distant source external to myself, namely star X." If he is a scientific astronomer, he will also make sure that his colleagues can endorse what he says.

Astronomers are chiefly interested in messages which, they all agree, are transmitted to them from heavenly bodies. Geologists are similarly interested in messages which, by common consent, are transmitted by rocks. And so on with other scientists. Psychologists are interested in messages transmitted by human beings, or possibly by other animals, and thereby are respectably in the company of scientists. But neurologists also are interested in human messages, and neurologists all agree that they are specially interested in messages which are referrable to physiosomatic happenings in the human body.

Now every human message is a physiosomatic happening, whether it be in process of transmission in the nervous system, or in process of being signaled vocally or gesturally. Psychologists are not specifically interested in messages as physiosomatic happenings, though they are not disinterested in them as such. In which messages, then, are psychologists specifically interested? They are specially interested in those human signs and signals which convey information or messages about the way in which a human being deciphers and relays other messages which he, in turn, has received. They are interested in the process of information and communication itself. One may reasonably say that psychologists are interested in mind, if one

defines mind as the information and communication function of the human organism. Since information and communication are intrinsically bound up with physiosomatic happenings, the concepts of information, communication, messages, signs, signals, and decipherment are eminently suitable ones for circumventing the bedevilment which ensues when mind and body are conceptually separated.

One way of doubly ensuring that outmoded connotations of separateness between mind and body do not sneak in through the back door is to avoid using the terms mind and body altogether. For this reason, the word ego may be used in place of any combination like bodymind. The term ego is a term that investigators may eventually agree upon as the one to use when they refer to the information and communication function of the human organism, as contrasted, say, with the respiratory function, the circulatory or digestive function, and so forth.

Like the solar system, the ego is a going concern. It operates all the time. Until the concept of gravitation was formulated, it was considered necessary to attribute local forces of attraction and repulsion to each of the planets in order to explain the functioning of the planetary system. Likewise in chemistry, before oxygen was discovered, combustible substances were said to possess phlogiston which caused them to burn. So also with the ego: it has been considered necessary to find sources of force or power or motivation in order to explain the functioning of the ego. It is quite sufficient, however, simply to take for granted that the ego operates or is functional to begin with. It does not need to be set in motion or made dynamic, for it does not stop. It is a going concern.

An investigator of ego begins his investigation by taking note of the signs and signals that a human being—possibly an animal—makes or displays. These signs and signals convey to the investigator some information about the way in which the subject has himself deciphered or interpreted yet other signs and signals that came his way. Sometimes the subject's signs and signals are simply verbal reports of the messages he deciphered. At other times they are gestural intimations of the messages he deciphered.

A point about gestural signals is that they, rather more lucidly than verbal signals, are instructive as an example of how signals do double duty. Gestural signals convey also that a person's decipherment of signs includes a strategy or plan of action. Thus, a soldier deciphers signs and signals to mean that an enemy sniper is in a tree to the left; the soldier takes aim and fires at the sniper. To an observer, the soldier's gestures of aiming and firing are a message that the soldier has deciphered a sniper in the tree. But the soldier's gestural signs also inform the observer that the soldier's decipherment shows itself as a strategy for solving the quandary of the sniper-in-tree signs. So it is that a given set of human signals may tell not only that a person has received a certain piece of information, but also that he is carrying out an operation of mastery in response to the challenge of that information.

To continue the military example: a second soldier also deciphers signs and signals to mean that an enemy sniper is in a tree to the left: he falls behind a hump of earth, immobilized. This time the soldier's gestural signs inform an observer that the soldier's decipherment of the sniper-in-tree signs showed itself as a strategy—albeit unpremeditated—of collapse, and of constraint instead of headlong mastery.

The investigator of ego can group the information he obtains from a given sample of his subject's signs and signals into three categories. In the first category is information pertaining to the subject's own decipherment and appraisal of signs and signals, whether their source be proprioceptive or exteroceptive. In the second category is information pertaining to the subject's mastery over yielding and collapse in connection with the resolution of perplexities and quandaries presented by various signs and signals. In the third category is information pertaining to the subject's control of mastery and attack in connection with the resolution of perplexities and quandaries presented by various signs and signals.

The investigator of ego is not logically

irresponsible if he goes one step further and subdivides, on the basis of these three classes of obtainable information, the function of the ego into three subsidiary functions. These three subsidiary functions may then be called spectatorship, mastery, and control, respectively.

From a sample, or series of samples of the signs displayed and the signals exhibited by a person, an investigator obtains some information about the operation of spectatorship, of mastery, and of control, all three, in that person. Over a given period of time, the three do not operate with equal salience, but one is ascendant over the other two. For example during sleep control is the ascendant and mastery in abeyance. Spectatorship is somewhat in abeyance, but from time to time dream signals or even perceptual signals, may get through and be deciphered as dream messages. The inhibitive control of sleep may effectively operate against the interruption of these dream messages. If the messages are strongly interruptive, however, the inhibition of sleep may be overcome by the stirrings of mastery, and the person wakes up. Mastery may then come still more into operation, as would be the case if the sleeper waked shivering, to find all the blankets on the floor.

Spectatorship, mastery, and control do not operate independently of one another, but together in closely interlinked intimacy. The operation of any one of them is a contributory influence on the operation of the other two, either by increasing or decreasing the degree to which the other two operate. In this respect, the operations of spectatorship, mastery, and control may be likened to the operation of a feedback system in engineering. To take an analogy from engineering: three ingredients A, B, and C are poured down three separate chutes into a single vat to form a mixture in the ratio of 5:3:1. As soon as too much of A gets into the mixture, a signal goes to the outlet of chute A to decrease its aperture, and a signal to each of B and C to increase their apertures, and so on, until the correct ratio is obtained. The quantity of each ingredient coming into the vat regulates the inflow of each of the other two ingredients.

The drawback to this analogy is that in the interlinked operation of the three subsidiary functions of the ego, spectatorship, mastery, and control, there is no fixed and predetermined ratio of their participation. It is as if, in the analogy above, the formula for the mixture were being constantly changed. Mastery is to the fore when a man rushes to get to work in time in the morning, for example. Then control builds up to a great increase as he restrains himself while the boss bawls him out for being late. He settles to his desk and spectatorship predominates as a vast array of new bookwork signs get deciphered, and so on.

It is the chemical engineer who changes the formula for the compounding of a chemical mixture. Who or what is the agent that dictates the changes of formula for the mixture of the three subsidiary ego functions? To this question various answers are extant. First, there is the answer that it is all a matter of free will—of personal selection and choice. Second, there is an answer in terms of biological constitution and hereditary determinants. A third answer is a stimulus-response hypothesis of social and environmental determinism. A fourth answer attributes causality to the impact of life-history experiences as the determinants, perhaps unconscious, of a person's subsequent attitudes, dispositions, or reaction tendencies.

None of these four answers is manifestly absurd, but each by itself is too restricted. Moreover, the four are neither mutually exclusive nor, as a group, exhaustively complete. They pose more theoretical dilemmas than they resolve.

There is another, a fifth answer, which does not pose theoretical dilemmas. It is this: the agent that determines changes in the formula for the mixture of the three subsidiary ego functions is not, strictly speaking, an agent at all. Rather, these changes are determined by a host of agents, operative over the course of a single life history, that cannot be well enough demarcated and distinguished from one another to permit simple statements of cause and effect. Like the Brownian movement of subatomic par-

ticles, this multiplicity of agents needs to be conceived of according to the language of the laws of probability or chance.

In the ultimate analysis, then, changes in the formula for the intermixture of the three subsidiary ego functions are dictated by the combined impact of a multitude of factors which, to all intents and purposes, operate as probability or chance. To postulate this proposition as an axiom is not to deny the possibility of scientific endeavor with regard to ego functioning. Use of the concept of probability in nuclear physics did not impede, but facilitated scientific endeavor.

Use of the concept of probability in investigations of ego is simply an admission, at the outset, that it is neither expedient nor pertinent for an investigator to be concerned with concepts of absolute or ultimate causes. He may then proceed to describe what he observes, to describe regularities of sequence in the happenings he observes, and to search for methods of intervention which predictably change what he observes. His interventions may be manipulations of genes and chromosomes, as in animal breeding and radiation experiments. They may be manipulations of a biochemical nature. Or they may be manipulations of the light and sound waves a person receives, that is, of the auditory and visual messages and information he becomes acquainted with. And so on.

Constant changing of the formula for the admixture of spectatorship, control, and mastery as the ego operates sequentially in time, and in changing situations, makes it improper to conceive of an ideal of ego-functioning as one in which there is an equilibrium between spectatorship, mastery, and control. One may legitimately conceive of an ideal of ego functioning, however, as one in which neither spectatorship, mastery, nor control become chronically and persistently salient. When any one of these three subsidiary functions operates too persistently and the other two are in abeyance, then the ego is evaluated as disordered or nonhealthy in its functioning. The criminal, for example, shows signs of chronicity of assaultive mastery. The depressed patient displays a chronicity of inhibitive control. The hallucinating schizophrenic patient gives evidence of a chronicity of spectatorship. To make such statements as the three foregoing, let it be pointed out, is not to offer causal explanation, but simply description of what happens.

Evaluation of the entirety of ego functioning on the healthy-nonhealthy scale is greatly facilitated if details of context and circumstances are specified for particular ego operations. Thus, control functioning may become excessive in a child only when the father is at home, or mastery may go awry only after a boy goes to the city to live, and so on.

Among the multitude of circumstances in which the ego operates, some classificatory order may be introduced by application of the principle that, for every member of the human species, there are some things about being human which are inevitable. Five broad categories of the inevitables of being human can be distinguished.

1. **Aggregation.** The human species is aggregative. A person is a unit in an aggregation, namely of the mother-child pair, when he is born. Only in the course of exceptional circumstances is he not a member of several other aggregations during the course of a lifetime. Some sort of a family or kinship aggregation is ethnically universal. A person responsively encounters other people and initiatorily transacts with them in a sort of social symbiosis—in a social interplay of positive and negative, give and take.

2. **Ecologic setting.** The human species finds itself in an ecologic setting which includes animals, plants and nonorganic stuffs and things. A person is an ecologic unit that impinges on, and is impinged on by other units and elements of the ecologic entirety. Growing older and maturing is intimately associated with the extension of the range of encounters and transactions with other ecological units, particularly with reference to food, shelter, and clothing.

3. **Subspecies status.** The human species can be construed, to use a piquant figure of speech, as carcass-things that function. In these carcass-things, several functional parts may be distinguished, namely the various parts or organs of the body. Distributed

among the human species, there are differences among parts of the body and their function that permit divers subspecies classifications. People do make these classifications, though not always systematically. Thus, there are "colored" and "white"; obese overnourished and emaciated undernourished people; hyperthyroid, euthyroid, and hypothyroid people; enuretics and nonenuretics; people with big noses and people with little noses, and so on. One single member of a human subspecies does, on occasion, encounter his bodily parts and functions and he carries out transactions on and with them as if the body were a carcass-thing independent of and separate from self. Thus, he says that so and so was done by instinct or reflex, that his skin color caused him to lose his job, or that his congenital deformity made him irritable and depressed. He cuts his fingernails, shaves his beard, and he may mutilate his own body sacrifically or cosmetically, and so forth.

4. **Sexual status.** A person is at birth equipped eventually to be either a mother or a father, or in the rare case of human hermaphroditism, more a mother than a father or vice versa. Beginning with a male or a female genital morphology, a person thenceforth establishes his prereproductive and reproductive gender role and orientation as male or female in a myriad encounters and transactions.

5. **Death.** A person is a unit in a chronological setting at the termination of which his unity is disrupted. The termination may be hastened or threatened by sickness or injury. The vast imponderableness of termination is the archetype of all the perplexities that confront a person, and the one for which the great mythologies, religions and philosophies of mankind have sought an answer. A person has, from time to time, various encounters and transactions, directly and by proxy, with sickness, injury, and bereavement.

These factors (and possibly others) are, then, the immutables of the human condition. They may be expected to be among the variables determining in a given situation the relative preponderance of the components of ego-function. A scientific investigation of mind, therefore, must operate within the limits set by these variables. To the extent that the variables are amenable to treatment by standard scientific methods, it appears possible to work toward resolving the dilemma of mind-body dualism.

(Manuscript received March 19, 1956.)

The significance of man is that he is that part of the universe that asks the question, What is the significance of Man? He alone can stand apart imaginatively and, regarding himself and the universe in their eternal aspects, pronounce a judgment: The significance of man is that he is insignificant and is aware of it.

—Carl Lotus Becker in *Progress and power*

SOME DISFUNCTIONS OF INSTITUTIONAL IMBALANCES

by Melvin M. Tumin

Department of Economics and Sociology, Princeton University

In most societies, individuals perform in several roles associated with several institutions. The rewards for quality of performance may be given on different bases in the different situations. For example, rewards in the economic sphere may be in money and within the family in affection. However, in some societies, as in our own, one type of role (and the associated institutions) may become so important that its values "invade" other roles and institutions. As a result, the evaluations of performance proper in the dominant institutions tend to be substituted for evaluations proper in others. The results may be strains on individuals and on the over-all goal achievement of the society.

In most if not all human societies one of the several institutions essential for social viability is given special emphasis. There is also a corollary attention to the means by which the functions of this institution are achieved and upon the values inherent in the means-ends sequence. Sometimes religious devotion is stressed; some societies are highly political in ultimate orientation; others emphasize primarily the facts and fictions of kinship; and in still others, the production of goods and services assumes the highest importance. Which institutional-complex is emphasized can vary over time within any single society.

Any set of institutions can coexist such that, as the human actors perform the tasks called for, by the means prescribed, within the limits of the values affirmed, they may find themselves impeded in their pursuit by contradictions within the separate institutions; or they may find that their activities in each institution are mutually facilitated; or what they do under one institutional aegis may have few consequences for the achievement of other institutional ends.

Emphasis on one institution may produce certain strains and stresses which affect behavior in other institutional areas. The amount of stress and strain will vary with the character of the emphasis. For instance, if most available social goods and services are distributed as rewards for performance of roles in the emphasized institution, other institutions will suffer from absence of rewards required to motivate performance of necessary roles.

Considerations like these indicate the effects of alternative types of institutional emphases. One may probably account for varying intensities of stratification in a society in terms of the relative emphases given to various institutions. This paper focuses upon one particular emphasis, that given to economic institutions in modern United States. How does the economic structure—especially job-recruitment, placement, performance, and reward—compare with the structure of kinship-role-playing, and how do these structures interact? Among the distinguishing elements of the economic structure are the following:

1. There are formal mechanisms for identifying, recruiting, training, and allocating talent to its most appropriate niches.

2. There is a specialized division of labor, organized in terms of a hierarchy of skills.

3. There are formal measures of role efficiency, formally stated and publicly known norms of performance, and formal mechanisms to control deviation from these expected norms.

4. There is a system of rewards by which different positions in the skill hierarchy are differently rewarded.

5. The rewards are measured in units which can be compared.

6. They consist primarily of scarce and desired goods and services.

7. The importance to the society of a given occupation is measured by its location in the skill hierarchy.

8. Prestige is allocated to a position proportionate to its place on the ladder, and, as a corollary, different positions are considered to have different moral worth.

9. Rewards and opportunities not immediately associated with prestige are also distributed according to place in the hierarchy.

By contrast, the familial institution has the following elements:

1. There are few if any formal mechanisms for location, recruitment, and placement of differential talents, and indeed the matter of talent is virtually not raised.

2. The division of labor is rudimentary. Sex and age criteria are employed in role differentiation, but without any formal or implied hierarchy of skills.

3. There is relatively loose definition of what constitutes efficient role playing, though there are norms which prescribe the range of acceptable performance. At the lower end, there are legal definitions of such things as neglect and desertion; at the upper end there are only informal community restraints against overperformance.

4. There is no formal system of rewards.

5. Such rewards as are present are not formulable in measurable units and hence cannot easily be compared from one position to another.

6. The rewards primarily consist of the gratifications inherent in the role-playing itself and such community approbation as one can command, and, at least in theory, these rewards are abundant rather than scarce.

7. All the role players who stay within the accepted normative range are considered equally important to the society. No hierarchical ladder of greater and lesser functional value can systematically be constructed.

8. There being no hierarchical ladder of importance, it is not possible to assign differential prestige for differential position. If there are any role-evaluations, they are made on the basis of the conscientiousness with which one fulfils his expectations, given his resources.

9. There are no distinctive rewards or opportunities arising from differences in role-performance.

When we juxtapose the foregoing two sets of features, it becomes apparent that there are genuine differences between the economic and kinship institutions, and that at least some aspects are mutually exclusive.

Traditionally we have viewed the operative economic rules as deviations from highest principles of conduct, but have also tended to assume that much of this deviation is unavoidable. Similarly, we tend to look upon familial role-playing as embodying a set of principles which are more in harmony with our highest values, but which are capable of being followed only within the family province. In short, the ends desired from economic and familial effort and the means required by those ends are considered to be different enough to warrant disparate organizations and norms.

Forms of invasion of family life

In actual fact, we note an increasing tendency for certain economic rules of conduct to invade the family sphere, coloring the quality of family life.

The invasion assumes a variety of forms. These include:

a. Competition among parents for invidious distinction as parents in terms of the goods and services with which children are supplied.

b. Judgment by the wife of the adequacy of the husband in terms of the economic level of his "breadwinning" and the prestige of his occupation.

c. Awarding of affection to children on the basis of achievement.

d. Emulation by families of a style of life whose worth is measured by economic resources.

e. Judgment by children of the adequacy of their parents in terms of the economic level maintained.

f. Segregation of clusters of families from each other, residentially, socially, and edu-

cationally, on the basis of economic-class attributes.

These tendencies may be identified as imports from the economic sphere, insofar as they derive from the occupational structure, and the ways in which prestige-allocation is tied to income and occupation.

Consequences of invasion

The entrance into family life of economic norms and values has consequences not only within family structure, but also, resonantly, in economic and general social structure, as well. For example:

1. *Production of dissensus through class differentiation.* If consensus in the society on major values is important and if the family is the chief agency for developing such values, then, consensus is hampered relatively to the extent to which family units are set apart from each other, and placed in different relationships to the norms, by virtue of differentiated economic position.

2. *Reinforcement of psychic strain.* If psychic stability is valued; and if the job situation, by virtue of its competitive and impersonal atmosphere, generates psychic stresses; and if personalized and individuated acceptance and affection in the home helps restore equilibrium; then it follows that altering this home atmosphere in the direction of a market-place orientation will maintain and reinforce rather than diminish the psychic strain generated at the job.

3. *Threats to social cohesion.* The cohesiveness of any society significantly depends on the extent to which its members sense their value to their society, and in turn value their membership. The economic structure operates so that differentiated rewards in power and property also yield different amounts of prestige or units of social worth. By contrast, the structure of family life is such that any parent is considered equally worthy with all other parents if he plays his parental role equally conscientiously. Those who rate highly on the economic index of social worth will tend to generalize this rating to an overall social position. And comparably those who rate less well economically, but who can esteem themselves for their conscientious fulfilment of parental duties, will tend to generalize their self-images derived from their parental roles to a more general social image of themselves. In short, favorable self-definitions achieved through conscientious parenthood help to balance unfavorable self-images acquired in devalued economic roles. Conversely, though perhaps questionably and in fewer cases, economic success may counterbalance failures in family relationships. So, multiple institutional affiliations provide opportunities for larger numbers to develop a sense of worthwhileness than would otherwise be possible, and this results in their feeling more valuable to society and so prizing more highly their membership in it. Therefore, any influx into family life of invidious distinctions based on economic criteria is likely to diminish social cohesiveness.

4. *Impact on personality development in children.* A major motif in family life, which is claimed to be systematically important for adequate personality development, is that no discriminations be made with regard to the amount of affection given to various children by their parents. Such a non-discriminating system is possible only where all children are considered to have equal rights to this affection by virtue of their being children of the same parents. One observable effect of the invasion of economic norms is the tendency to introduce "achievement" as a criterion by which to judge the amount of affection a child is to receive, and, in accordance, to award different amounts of affection in proportion to differential achievement. If the original premise regarding the need for equal affection as a basis for sound personality development is valid, it follows that the introduction of "differentiated achievement" may impair the family's role in personality development.

5. *Dangers for job productivity.* Productivity on the job varies relatively to the extent to which the producer feels his productive role is valued. The differentiated reward structure of economic activity distributes this sense of valued-role unequally. If such distinctions enter into other institutions like the family, they may make for more generalized discontent which can ultimately lower

productivity in the economic sphere as well as impair the functions of the family.

6. *Impact upon family stability.* Since the modern family has lost many functions it once had, the affectional tie is now the principal binder among the members. If, as a result of the introduction of economic norms, spouse relations are made to vary with the adequacy of the "breadwinning" of the male, then varying economic fortunes over time and differences between breadwinners at any one time provide two major sources of instability in spouse relations. The tendency to judge the adequacy of a husband by the income he is able to provide may help account for the now-evident greater instability of marriages in lower socioeconomic levels.

Here, then, are six ways in which the entrance into family life of certain economic norms and values may affect achievement of economic, familial, and more general social goals and values.

The consequences of concentration and unequal distribution of rewards

What are the causes for the spread of normative principles across institutional lines? To what extent does the system of distribution of rewards set the process in motion?

The line of events can be spelled out as follows. Let us assume, to start, that any society has to offer, by way of rewards, three items: property, power, and prestige. These can be distributed equally or unequally, and they can be distributed primarily within one institutional framework (e.g., as job rewards alone, or religious rewards alone), or they can be spread through various institutions (e.g., property rewards for the job, prestige rewards for religious purity.)

Four major types of distribution-systems are thus made possible by the variations inherent in the possible combinations: (1) equal-concentrated; (2) unequal-concentrated; (3) equal-diffused; (4) unequal-diffused.

1. The equal-concentrated system is one in which major social rewards are distributed primarily for role performance in a limited number of institutions, in which the tendency is toward equality of reward because "conscientiousness at role" is the criterion of merit.

2. The unequal-concentrated system is one in which, as above, the rewards are distributed in only a few situational frameworks, but here the tendency is toward maximum inequality, because "differential talent" or "differential birth" are criteria of merit.

3. The equal-diffused system is one in which various desired scarce goods and services are allocated to reward performances in a range of institutions, and in which the tendency is toward equality of reward as noted in (1) above.

4. The unequal-diffused system is one in which as in (3) above, rewards are spread through various institutions, and in which the tendency is toward inequality as noted in (2) above.

If we take "adequate role performance" as our end, and if we assume that such performance requires sufficient reward as motivation, then two hypotheses are suggested:

(1) The fewer the institutions within which such rewards are adequate, the lower is the level of role performance likely to be within those institutions which have inadequate rewards. Conversely, the more institutions there are with adequate rewards, the higher is the over-all average of role-performance likely to be, though, in any given institution, the performance may be below that which it would be if it had a disproportionate share of the rewards.

(2) The more equal the chance for everyone to gain an equal amount of the available rewards, the higher is the over-all average of motivation likely to be in any given institution, and throughout all institutions, though the motivation of some individuals may be lessened because there are no distinctions between themselves and their colleagues. Conversely, the less equal the chances for equal reward, the lower is the over-all average of motivation likely to be, even though the motivation of some individuals may be higher.

If these hypotheses prove valid, then it follows that the unequal-concentrated system of rewards is likely to yield the lowest over-all average motivation for adequate

role performance, and the equal-diffused system would likely yield the highest overall average.

It is further suggested here that probably the equal-concentrated system would be more generally productive than the unequal-diffused system. The premise here is that the importance of concentration of rewards is diminished by the fact of equality more than is the importance of inequality diminished by the fact of diffusion.

The institutional arrangement in modern United States resembles most closely the unequal-concentrated type. While it is not possible empirically to demonstrate that alternative arrangements would yield a higher average motivation to conscientious role performance, we can suggest ways in which the fact of concentration and inequality of distribution of rewards operate to produce the effects previously cited. Alternative arrangements can then be compared in terms of their likely effects.

Two major effects of the American institutional arrangement are evident. The first is that as a result of the great emphasis placed upon job-achievement and its correlates, it becomes difficult to give enough socially-rewarding meaning to the values of other institutions to enable them to compete successfully on their own terms. There is relatively inadequate reward for conscientious role performance in other institutions. As a result, values from the economic institution invade other institutions and we get the consequences which we have already briefly outlined. This invasion is disfunctional to the extent that the entering values are inappropriate to the roles to which they become attached.

A second effect is the strain created by the presence of publicly-symbolized inequality in fundamental conditions of life in a society which places a central value upon the equality of worth of all men. In a caste-society, where acceptance of birth-inequalities is the rule, these differences are not likely to be of great significance. There is no presumption of equal opportunity, and hence no implied moral condemnation for relative failure. But in modern American society, the notion that equal opportunity is available is built into the core of economic ideology. Hence such differences as appear are taken to indicate greater or lesser adequacy, importance, and ultimately, moral worth. On the assumption that all men strive to maximize the number of favorable definitions of themselves in their effective environment, we ought to expect that men will react strongly to the implications of unequal worth embodied in the unequal distributions of the symbols of worth, namely, property, power, and prestige. By and large these reactions ought to take the form of behavior designed in some way to reduce the inequalities, to diminish comparisons with others, or to deny the appropriateness of the criteria of evaluation.

The scope of these problems varies with the range of statuses to which one is urged to refer himself for comparison.

In a caste society the problems generated by inequality are likely to be confined within castes since relative self-evaluation tends to be intra-caste. In an open class society, however, the problems are likely to assume society-wide proportions since relative self-evaluation tends to be inter-class. In short, inequality is likely to create problems wherever it occurs, but its scope will vary with the system of stratification.

Reactions to symbolized inequality

It is not possible at this point to outline systematically what sorts of compensatory reactions are pursued by individuals of various personality types. We do not yet know enough about the differential incidence of these reactions to construct adequate typologies. But it is evident that reactions directed toward compensation for publicly symbolized inferiority are widespread. A number of them may be listed here:

1. Spending disproportionate amounts of one's resources on status symbols which help confuse the status picture and give an appearance of higher standing than the individual in fact enjoys.

2. Rejection of the values of this world and immersion of self in a belief system which defines earthly inequalities as temporal and insignificant and other-worldly equalities as eternal and prime.

3. Compulsive denial of the significance of

the publicly accepted criteria of worth, and insistence upon the equality of all men as measured by criteria which cannot be publicly symbolized. Often associated with this orientation is spontaneous and aggressive reaction to any suggestion of inequality, resulting sometimes in belligerent public conduct to prove to oneself and to others that "no one can push me around."

4. Excessive reliance upon sentimentality and warmth, as criteria of the worthiness of others; the cult of friendship; the constant search for public affirmation that one is loved.

5. Exaggerated authoritarian behavior in those situations where power can be exercised, such as in the home.

6. Denial of the importance of distinction for one's self, together with investment of one's hopes and aspirations in one's children, as a means to become elevated, by ascription, to a status one could not reach by personal achievement.

7. Rejection of existing criteria of worth as immoral, and participating in political movements whose aims are nominally to establish a society in which more valid criteria of human worth are dominant.

8. Confining oneself as much as possible to interaction only with one's peers, thus reducing the effectiveness of symbolic reminders of inequality.

It must be asserted immediately that not every incidence of these tendencies is to be interpreted as compensatory behavior. We recognize the *sui generis* legitimacy of many of these tendencies. And it must also be admitted that all these reactions can occur within the range of socially accepted norms. But we also recognize that each of these reaction types is capable, without much extension, of developing into a normatively disapproved, if not psychiatrically pathological, mode of conduct.

If, then, we can safely identify these as among the common ways in which persons react to a sense of their inadequacies in property, power, and prestige; and if, in turn, we can identify these inadequacies as resulting from the fact that the major rewards of the society are distributed unequally, then, it would seem efficient, in order to maximize goal-achievement, to make it more possible for larger numbers to achieve more nearly equal rewards. This assumes of course that we desire to continue achieving not only our economic goals but also those results from family life which we can realize only if it is not impinged on by inappropriate values from the economic sphere.

We have provisionally identified the concentration of rewards within one institution as a major source of social imbalance. The natural suggestion is, of course, to bring about greater diffusion. But this involves complex problems regarding the dominance and subordinacy of various institutions within a society. Though every society emphasizes certain roles as more important than others, the identifiable problems arising from *great* concentration would suggest that every society has a vested interest in being cautious in permitting any one institution to become preponderant.

Sources of institutional dominance

In this regard, a number of ways in which institutions may become and be dominant can be cited:

1. Through monopolization of the scarce and desired rewards.

2. Through excessive demands upon time and energy.

3. Through generalized social emphasis upon the products of the institution.

4. By virtue of the strategic character of the products or services, as when no activity can be undertaken without ritual by specialists.

5. By virtue of ignorance of alternative possibilities.

This is a partial check list of the kinds of imbalances of which any society must ever remain wary if it is to prevent consequences which hamper attainment of a wide range of desired goals.

(Manuscript received December 24, 1955.)

A CRITIQUE OF P. E. MEEHL'S *CLINICAL VERSUS STATISTICAL PREDICTION*

by Richard D. Mann

Mental Health Research Institute, University of Michigan

It is important both for the actuary and for the clinician to be able to predict certain events. The actuary in making predictions uses straightforward, precisely describable rules, so that in principle anyone able to follow these rules (a clerk or even a machine) can make the same predictions. The question has been raised whether in the clinical situation similar methods of prediction should be introduced or, indeed, whether the clinician already "unconsciously" uses similar methods. The controversy surrounding this issue has been intense. Meehl puts the focal points of the controversy under close scrutiny.

THE intramural battle that has been waged in all the social sciences, but particularly in psychology, over the proper role of human judgment in the prediction of future events is neatly introduced by P. E. Meehl in his book *Clinical versus Statistical Prediction*.[1] In the first chapter he lists over eighty adjectives, some honorific and some pejorative, which the contestants have used to describe themselves and each other throughout the long debate. As a catharsis, the list seems to have served its purpose for the author, since the remainder of the book is a calm and searching investigation of the issues. But the list leaves no doubt in the reader's mind that the distinction implied by the "versus" in the title has been a matter of grave concern to social scientists of all persuasions.

This volume is a collection of lectures given at various times and places, but taken together they cover a wide range of the issues in the debate. The first audiences for these papers were composed largely of clinical psychologists, it might be guessed, but the problems raised here have generality far beyond situations in which people are called upon to make predictions. Although Meehl speaks of the clinician, and we will follow this orientation in the review, the arguments are addressed to all who attempt, in the face of varying degrees of uncertainty, to assess the outcome of a course of events.

An outstanding characteristic of the book is the author's ability to subdivide the problem into manageable segments, to set up comparisons between the two approaches in relation to a large number of logically separate questions. A major part of the book is an exposition of the clinician's activities: first, at the stage when he is confronted with the observable world, next at the stage when he enters the world of science by attempting to communicate a prediction, and finally when he attempts to evaluate his procedures with regard to their accuracy, efficacy, and efficiency. The author spends a considerable amount of time seeking to find out why and at what point in the total process the two techniques diverge; only then does he move to consider the logical as well as the empirical grounds for maintaining that one is more satisfactory than the other.

The author's first discussion distinguishes between the kinds of data used (either psychometric or not) and between the methods

[1] *Clinical versus Statistical Prediction, a Theoretical Analysis and a Review of the Evidence* by P. E. Meehl. Minneapolis, Minnesota: University of Minnesota Press, 1954. Pp. x + 149.

used to combine the data (either mechanical or not). Psychometric data are defined by four properties: standardized methods of collection, immediate recording, objective methods of classifying the response, and the existence of norms. Mechanical combination consists of the application of explicit, computational rules—as in the multiple-regression equation. Furthermore, a prediction may not only combine psychometric and nonpsychometric data, but may involve application of mechanical rules of combination to *scores* which are themselves the product of nonexplicit human judgments. Thus, a prediction must be classified as statistical or not in relation to a particular level of interest. One distinction between the actuarial and clinical approaches lies in their use of mechanical and nonmechanical methods, respectively.

Meehl turns next to the question of the reasonableness of predicting from class membership. He cites Allport's seeming distrust of those statements about an individual's likelihood of performing a certain act which are based on considering the individual as a member of a class of individuals. In reply, the author points out that the classes could be based not only on sets of individuals categorized according to a common characteristic, but also on the probability that an act will occur in a given situation or the probability that a particular individual would perform such an act in any situation. Carrying this argument further, the point is made that while the individuals themselves may be the basis for a set of relative frequencies, the optimum subclass for a particular actuarial prediction may be the class representing the interaction of the single individual in this one situation. Prediction from class membership is not the same as "equating" a set of individuals, nor does it necessarily mean "losing the uniqueness of the individual."

On this point Meehl follows Reichenbach who states that there is no such thing as one single probability of an event, rather only a series of probabilities dependent upon classes which are smaller and smaller extensionally and more and more complex intensionally, until finally the number of observations in the class becomes too small to provide stable probability estimates. The optimal subclass is the smallest class of the sequence sufficiently large to provide such estimates.

The author's next concern is to describe just what there is about the clinician's brand of prediction that distinguishes it from that of the statistical clerk. For one thing, he can use a different set of "beta weights" in the regression equation; for another, he can use different sets of predictor variables. Although he recognizes that those who use and defend the actuarial approach would argue that if a particular fact were relevant it should be used by the clerk, Meehl states that in the real world of clinicians and clerks it is likely that certain rare events which do in fact have relevance for a prediction would not have been studied sufficiently to allow the clerk to make use of knowledge about their occurrence. Meehl notes that the question of whether introduction of these rare events into the prediction will increase its accuracy is another issue, but he points out that the power of the human brain in certain perceptual capacities may create more sensitive analysis of the person that present-day descriptive statistics will allow. The isolation of patterns among the data would be the best example; a clinician handling the Rorschach may be said to make internal calculations of greater complexity than the clerk usually deals with.

Perhaps an example would be helpful here. Meehl cites a report by Reik of a postdiction under rather unique conditions. During an analytic session one of Reik's patients complained of a toothache, resulting from the extraction of a wisdom tooth. After a long silence the patient pointed to the bookcase and remarked that there was a book standing on its head. Reik reports, "Without the slightest hesitation and in a reproachful voice I said, 'But why did you not tell me that you had had an abortion?' " Meehl's point is that the clinician is able not only to use cues so rare as to be very unlikely to occur in the actuarial table but also provide the theoretical sophistication necessary to place the cues in larger classes. It makes interesting speculation to wonder

just what one would tell the coding clerk to do with such cues as the reported tooth extraction, the silences, the remark about the book on end, and so on. Reik's perception of the connection between the tooth extraction and the birth process, between the upside-down book and the fetal position, and between the silences and resistance certainly is not immediately obvious. The connection depends on a sensitive use of theory, as well as considerable experience. It is clear that a creative act has occurred in Reik's inference, and the statistical clerk would be hard pressed to compete at this stage of the process unless, as Meehl points out, he were already a sensitive clinician. Still, even in this situation, it is possible to argue that the "actuarial" process occurs. The point is that in the mind of the clinician the "unique" situation *is* lumped with other equally "unique" situations which have some characteristic in common. Thus, the clinician may define a class of situations all of whose members are particular symbolizations of abortion. It is in his ability to define (or create) such classes (perhaps unconsciously) that the clinician has the advantage over the clerk, even though the estimation of probabilities after the classes have been defined can in principle be an essentially statistical process.

In the course of making his own position clear, Meehl analyzes the arguments of T. R. Sarbin, a staunch proponent of the actuarial approach. Sarbin argues that all statements about events must be viewed as expressions of probability to some degree, not as certainties. This would apply equally to statements about single events and about mass phenomena. Each statement is confirmed, therefore, only if this implicit probability can be determined, and this can be accomplished only by placing it within a reference class. This postulate of an underlying frequency distribution constitutes for Sarbin our only reasonable hope for confirming or disconfirming any prediction. Sarbin goes on to state that the clinician and the actuary are going about the job of making predictions in the same manner, except that the actuary uses cues that are more identifiable and rules that are more explicit.

In reply, Meehl separates his analysis of how the clinician goes about getting to the prediction and how the clinician or anyone else would determine the accuracy of the product. He takes no issue with Sarbin's point that verification of the statement depends on ordering the event to a larger class, but he does not agree with the notion that the clinician is going through an actuarial process internally, at least not in the sense of the phrase he attributes to Sarbin. Meehl agrees with the clinician that statements about single events in the life of an individual are made not with reference to the individual as a member of a class. As Meehl sees it, it is only when one attempts to assign a degree of confidence to the statement that the criterion of confirmability becomes relevant. It is then that the reference class with its relative frequency should be used.

But even here Meehl attempts to distinguish between this assignment of confidence and the verification of the statement itself. He gives an example: a prediction that Jones will commit suicide within the year only takes a year of waiting before it is either confirmed or disconfirmed. On the other hand, if we assign some probability estimate to the prediction, as for example, "He has a fifty-fifty chance . . ." or "He is very likely to . . .", it is most certainly true that only a class of events similar to the one being predicted will help to establish its confirmation.

Meehl emphasizes that he would have no quarrel with Sarbin's statement that the clinician is proceeding actuarially if the term actuarial is interpreted in its broadest sense. Certainly, there is no question that the clinician, confronted with an array of facts about an individual, is going to depend on his past experience to arrive at a statement about the most probable future event for the individual. But the question is whether this process should be called actuarial in the sense implied by the actuarial table or the regression equation. In an attempt to answer these questions Meehl outlines what he feels is the role of nomothetic laws in the making of predictions. He would agree that a law is dependent upon the recognition of classes of events in the

observable world. But where is the logical connection, he asks, between the events and the most adequate, the most genotypic classes to be used in the prediction? The author is careful to point out that he does not question the deterministic truism that if we knew all the relevant factors, all the situations which the individual was about to face, and the individual's value on each relevant parameter that prediction would clearly be mechanical. The crucial role of the clinician, however, with the state of knowledge that obtains today, is first to define the classes with which to organize the data and second to decide the relation of the observed events to the hypothesized intervening variables. It falls upon the clinician to realize that a particular event is a specific case of the more general law involved. Thus, since the hypothesis is by no means uniquely given by the data at hand, the clinician is *forced* to the job of inductive reasoning if he is to make this bridge. On the other hand, the actuary is seen as performing a strictly deductive task, the prediction being a direct consequence of the known relative frequencies in the defining classes in combination with the established procedures for grinding out the probability statements. This basic difference, the creative act on the part of the clinician, is the basis for Meehl's statement that the clinician is by no means a poor substitute for the electronic computer; he is involved in a very different process, at least in certain stages.

In summary, Meehl maintains a consistent position with respect to clinical prediction. He takes issue with writers like Sarbin, not because they emphasize the importance of empirical confirmation, but because they insist that the clinician is doing the same things as the statistical clerk, only less explicitly and precisely. The clinician is faced with the job of reconstructing a person's total life history as accurately as he can. The observations he makes serve as "end terms," as leads to the individual's level on dimensions not observable, but only inferable. On the other hand, no matter how irrational the process of creating the hypothesis may seem, no matter whether the clinician is a psychologist or Reichenbach's clairvoyant, the criterion of confirmability will still be applicable. However, Meehl points out that the fact that the person doing the predicting is unable to identify the cues he has used is no indication that the process was irrational. The detective who is able to verbalize all the cues that enabled him to crack the case has few counterparts among clinicians. The difference between Sherlock Holmes and the practicing clinical psychologist on this point should be seen largely in terms of the strength of the habit to perceive *and report* the relevant cues. For a fuller treatment of the kinds of cues involved, the reader is referred to Sarbin and Taft's essay (1).

By far the largest single chapter in the book is a review of the empirical evidence, meagre though it is, which bears on the comparison between the respective accuracies of clinical and statistical predictions. The criterion for a directly relevant study is that the data used for prediction are combined mechanically (usually by a regression equation) as well as nonmechanically. Meehl provides careful criticism of the evidence presented and, in addition, includes some calculations of his own on the reported data. In summary he says, "In spite of the defects and ambiguities present, let me emphasize the brute fact that we have here, depending upon one's standards for admission as relevant, from 16 to 20 studies involving a comparison of clinical and actuarial methods, *in all but one of which the predictions made actuarially were either approximately equal or superior to those made by a clinician*" (author's italics).

Meehl's reaction to the evidence is interesting. He reminds the clinician that even if it is true that the clinician is doing something very different from the statistical clerk, he should not expect to be more accurate than the clerk. Sometimes, it seems, he will do as well; sometimes considerably worse. In his introduction, Meehl particularly stresses the point that if the two approaches are of equal predictive value, it might be well for the clinician to abandon his job as predictor and concentrate on his job as therapist. Questions often dragged through hours of staff conferences of who

would benefit from which form of therapy might then be handled by a statistical clerk, whose time, presumably, is less valuable.

One the other hand, the comparison of the predictive value of the two methods may not be relevant in all situations. A false interpretation in therapy may not necessarily lead to undesirable results. It may, for example, lead to new information about the patient's willingness to correct such errors. This is not to say, however, that anyone should ignore evaluations of the accuracy of his methods of prediction, particularly when predictions obtained by different methods disagree.

Throughout the book Meehl shows a great deal of tact in presenting both sides, but it does seem that he is a bit more generous with the clinicians. By this I mean two things: I do not feel that he records the sensitivity of such writers as Sarbin to the intuitive processes as fully as he might; secondly, it seems that he spends more time expanding on reasons, not usually given by clinicians themselves, why the clinical approach *could* be superior in accuracy to the actuarial than he does in spelling out fully the advantages of the actuarial approach. I suppose this tactic is justifiable because the actuarial approach, having been more concisely stated, does not call for a sympathetic interpretation to the extent that the clinical approach does. Actually, Meehl appears to be more concerned that the clinician understand the reasonableness of the actuary than vice versa. It would be a fair guess, at any rate, that the book was intended for and will be mostly read by those clinicians who, in Meehl's words, "In spite of possessing and respecting clinical know-how, nevertheless are committed to making the enterprise as scientific as possible." My only negative reaction to this form of presentation, however, is that Meehl does not give the supporters of the actuary sufficient credit for their awareness of the reasonableness of the clinician's views.

To be more specific, Meehl quotes Sarbin's older works on the subject of inference and prediction in psychology, but he does not adequately cover the 1952 publication by Sarbin and Taft (1). There, the position is taken that the clinician is proceeding by a series of inferences, and that these inferences are, "made on the basis of prior experience—and that wittingly or unwittingly, prior probabilities function in the formulation of valid posterior probabilities." Such a broad meaning of the word actuarial as Sarbin here intends is accurately picked up by Meehl, but it leads Meehl to state that it takes "all the sting out of Sarbin's argument." It certainly does, and it makes me wonder further if Sarbin cannot, therefore, be relieved from his role as "boogey-man." Nowhere in his latest work is there any implication that the clinician is a second-rate Hollerith machine. In fact, Sarbin's views seem to be much more adequately summarized in his own words: "Since, in practice, the psychologist's inferences are used by himself and others in the appraisal of others . . . it is no less urgent that we turn more of our attention to identifying, isolating, and communicating the essential features of the inferential process." Sarbin, as far as he goes, seems to be arguing much the same position as Meehl; credit goes to Meehl, however, for his contribution to our understanding of the clinician's role in the formulation of the relevant classes, the fruitful hypotheses, and the isolation of the rare and complex events with which he attempts to deal. I do not feel that Sarbin would be unwilling to grant the importance of all these functions, but his articles have not treated them as sensitively as Meehl's book does.

Let us now turn to another way of looking at the activity of the clinician. Both Meehl and Sarbin refer to the familiar tactic of predicting the most likely event. In rolling two dice the strategy of predicting seven every time will pay off handsomely over time. But is the clinician really faced with an analogous problem? Let us return to an example given by Gordon Allport and cited by Meehl. If it is known that 80 per cent of all delinquents from broken homes are recidivists, what does it really mean to say that the probability of this particular boy (a delinquent from a broken home) becoming a recidivist is .80? We can say it in terms of four out of five chances, but we know he

either is or else he isn't going to become a recidivist. We are not going to "run him through" life over and over, as we would an experimental animal through a maze.

The approach that might be useful here is that taken by Savage (2) when he talks of personal probability. He would speak of the delinquent and our guess about his "chances" in rather different terms. What we seem to be doing by putting a probability estimate like .80 on the prediction is expressing our "degree of belief," our *subjective* probability that this particular delinquent will become a recidivist. The notion of a relative frequency seems to appear at two points: the first is in the process whereby we arrive at the degree of belief, and the second is when we enter a number of our predictions into competition with the random number table to see whether we really should ask anyone to pay attention to our predictions. Sarbin's phrase that we are proceeding on, "the basis of prior experience," is clearly meant to imply we make use of classes that help us to make sense of the data at hand. But the fact that we make use of stored memory traces does not mean that what might be called the subjective relative frequencies bear any relation to the objective ones. But accurate or not, we can think of a person's degree of belief in relation to not only one prediction, but in relation to a set of parallel predictions about the same person. For example, if I knew that a boy was six years old and in the fourth grade, I could verbalize some approximation of my subjective probability estimate that he would obtain a Stanford-Binet IQ of over 100, of over 120, and so on. Similarly, for each point on the scale I could announce my degree of belief that this would be his score. Which one would I choose, if I had only one guess? If I had access to data on six-year-olds in the fourth grade, I might predict the mean of this sample. Without any such data I am left with a number of subjective probabilities. The rational man would pick the *most* likely in this situation, as he would in the dice game, but how about the clinician?

Savage has proposed that we look not only at the subjective probabilities but at the utility that each act has for the subject. If we are to understand why someone bets on this treatment rather than that, it is important to keep in mind that not only the probability of cure is relevant but that estimated losses and gains resulting from a decision must also be considered. If, for example, studies showed that this year 3 per cent of all seven-year-old boys were going to die from polio and that if a certain vaccine were used, none would die from polio but 3 per cent would die as a direct result of the vaccination, it seems a reasonable guess that the parents (and probably the doctors) would oppose substituting man-made death for natural death. The consequences of the act of leaving the boy to a more mysterious agent of death are obviously greatly different from the consequences of doctor-produced illness. But consider the case of the psychiatric staff learning that, if left alone, 53 per cent of all nonpsychotic patients would leave the hospital for good; on the other hand, the figures showed that application of electroshock, insulin therapy, nondirective counseling, and psychoanalysis singly or in combination produce a remission rate of 53 per cent. It is extremely unlikely in this situation that the psychiatrist would just do nothing, nor would society permit such an economic measure as replacing the present hospital administrators with ones who would discontinue all therapy. I intend to raise no criticism whatsoever of a value system which would generate these decisions, but it seems important to realize that the value (that is the utilities assigned to various outcomes) as well as our estimates of probabilities of these outcomes is going to determine our decisions (our "bets").

The only point that needs emphasis among these thoughts is that when we look at the total behavior of the human judge as he attempts to predict a future event, it is important to know about as many factors as possible which influence his final decision. It is to be wondered whether Meehl should not have labored a bit harder on the aspects of human prediction, such as the introduction of different utilities for the various predictions, which tends not only to decrease predictive accuracy but also detract from

the relevance of this accuracy in the making of decisions. This is true especially of the clinician. Here may be the hub of the discrepancy between the clinician's intuitive feeling that he has made a good bet and the frequently discouraging level of his objective performance.

Meehl's style is smooth and direct. An occasional sudden barb provides ample amusement for the sophisticated reader. Perhaps the most noticeable quality of the book, to those readers who have been involved in discussions over this issue, is the degree of balance the author maintains with respect to both positions. Meehl's own professional experience as both practicing clinician and practicing experimentalist has made him an interesting fusion of the two approaches and the result is a book, which distinguishes the issues cleanly and handles each with intensity. There is no award given to one or the other side; much as in a well-argued and well-prepared debate, the decision to call it a draw is not meant to imply that both sides were equally right at all points along the way.

REFERENCES

1. Sarbin, T. R., & Taft, R. *An essay on inference in the psychological sciences.* Berkeley, Calif.: Garden Library Press, 1952.
2. Savage, L. J. *The foundations of statistics.* New York: Wiley, 1954.

These thoughts are many of them crude and hasty; and if I were merely ambitious of acquiring some reputation in philosophy I ought to keep them by me till corrected and improved by time and farther experience. But since even short hints and imperfect experiments in any new branch of science, being communicated, have oftentimes a good effect, in exciting the attention of the ingenious to the subject, and so become the occasion of more exact disquisition and more complete discoveries; you are at liberty to communicate this paper to whom you please; it being of more importance that knowledge should increase than that your friend should be thought an accurate philosopher.

—Benjamin Franklin to Collinson

ROUND TABLE DISCUSSIONS ON BEHAVIOR THEORY

In this department we present edited segments from round table discussions concerning behavior theory. These discussions are largely informal and unrehearsed, and they illustrate more the activity of the development of such theory and the interplay among individuals and disciplines, rather than finished products. At points their value more likely resides in the liveliness of the interchanges rather than the accuracy of the conclusions.

The present discussion revolves around the question of whether it is possible or profitable to view a network of "orientations" in a social situation as a "system" in its own right, i.e., one which can be studied in the light of its inherent structure and dynamics.

Orientation Systems

Theory Seminar, March 21, 1956, Mental Health Research Institute, University of Michigan.

Participants, all associated with the University of Michigan, except as indicated: David Easton, political science, University of Chicago; Ralph W. Gerard, biology (neurophysiology), Mental Health Research Institute; Arnold Horowitz, psycholinguistics, Mental Health Research Institute; Donald G. Marquis, psychology, Department of Psychology; Fowler McCormick, visitor; James G. Miller, psychiatry and psychology, Mental Health Research Institute; Theodore M. Newcomb, social psychology, Departments of Sociology and Psychology; Anatol Rapoport, mathematical biology, Mental Health Research Institute; George Ruff, psychiatry, Mental Health Research Institute.

Newcomb: My loci of interest today will be persons, viewed as systems of orientations; and groups, viewed as systems of systems of orientations.[1] My concern is to raise certain more general questions about this kind of system.

In the first place, I should like to ask again (as I did before) about the "reality" status of systems of this kind, that is, systems of orientations which in turn are viewed as inferred states of organisms. They are inferred rather than directly observed states. First I would like to ask what gains are there, if any, from viewing persons and groups as systems of this kind? What gains are there on the one hand, for psychology or sociology? And on the other hand, what gains are there, if any, for system theory in general? I also have special kinds of problems which will perhaps be worth discussion; not so much because of intrinsic interest in the subject matter but because of their generalization potential. First, what about certain emergent properties at the group level? (I am using emergent not in any evaluative sense at all, but simply in the sense of properties which appear at one level which do not appear at another level.) Secondly, what about certain properties at the individual level which are presumably abandoned when you move up to the group level? And does one lose anything significant by virtue of the fact that one does have to abandon certain properties at the lower level when you move up to a higher one, etc.?

Let's start with the first question I men-

[1] "Orientation" refers to the organization of an individual's psychological processes which affects his subsequent behavior with regard to an object or class of objects, discriminable by him. The psychological processes of particular importance are the affective, or cathectic, and the cognitive-perceptual. Roughly speaking, an orientation is equivalent to an attitude in the technical sense.

tioned. This question has been raised by Miller. What can be said about the "reality" status of orientation systems, which are at best only inferences about the states of organisms. Insofar as we measure orientations, and within reasonable limits we can do something which we call measuring them, it can be demonstrated that there are fairly stable relationships among them, that they seem to change in an orderly manner, etc. So from one point of view they must correspond to something "real" or they wouldn't maintain these orderly relationships. But on the other hand, I think we have to say that they are only inferred and never directly measured. Miller, does this qualify them for your adjective "real"?

Miller: I have been trying to see just why in the last session I was appearing stubborn in order to get over it just as fast as I could. Apparently some wondered if I was denying the epistemological that nothing is "real" in the sense position that you don't see the "apple," but you observe only specific aspects of it—redness, roundness, etc. Well, I had no intention of disputing that—I agree. And I wish we could stop arguing on that point. It has occurred to me that a better way to state what I mean is that a real system is one where something happens out in the environment which alters a variable as perceived by one individual or several observers, on which you get consensual validation. A conceptual system, on the other hand, is one where nothing happens outside the individual in order to alter this variable, but rather it happens within the observer. And you can still have consensus. You can have consensus because the observers have had similar learning experiences in the past and because they have similar neural structures. Now, I don't know whether that will help answer your question.

Newcomb: I think it is equivalent to saying, "yes."

Miller: I think so, too, because something is happening out in the environment which alters independently of anything that is going on in the observer. Then I would say yes, that's what I mean by "real." I don't mean you can get to the essence of it.

Gerard: I suspect, what I would much rather have resolved than whether it is "real" or not, is whether it is an important distinction and why. It seems to me that the real issue that we want to settle is whether this is a sufficiently self-contained system, and not whether it is a sufficiently "real" one to be interesting.

Newcomb: Shall we move on to the question of self-containment, then?

Marquis: Well, have we now said that the class of non-real systems is zero? Or what?

Newcomb: I thought that what Miller was saying was that there have to be transactions across boundaries, there has to be input which changes the state of affairs.

Miller: Yes, the state of affairs in the observer has to change as a result of a transaction across his boundaries from the outside—from the thing he is observing. If that occurs, and it is clear that this is why the change occurred, then this is a real system. If, on the other hand, a change occurs within the observer as a result of transactions which are entirely within himself, then we are dealing with a conceptual system.

Easton: Why not eliminate terms like "real" and "concrete" and simply talk about observable events and nonobservable events, or directly observable and indirectly observable events? We are interested in systems that are observable or indirectly or directly observable.

Marquis: Miller, do you use the word "observer" to mean the researcher?

Miller: The scientist. I assume that you must get consensual validation; two or more observers must agree about a phenomenon for it to enter the realm of science.

McCormick: May I ask a question to be brought up to date? In what sense are you people using the words "within" and "without" the observer? What is "within" and what is "without"?

Miller: Personally, I am talking about his boundary.

McCormick: What is the boundary?

Miller: His skin, his sense organs.

McCormick: Well, that would be the physiological boundary, but what would be the subjective in and out?

Miller: I don't think we make a distinction. At least, I don't.

Gerard: The minute you introduce consensual responses you are eliminating the purely subjective.

McCormick: Yes, I see the point.

Easton: I thought the issue was raised less with regard to conceptual systems like, let's say, mathematics, or Hegel's philosophy, than with regard to "action systems." I think of "real" and "non-real" as used in relation to action systems as against the whole acting individual. If we used the notion of "observable system," then, since actions are observable and also whole acting individuals are observable, we may say simply that there may be two different kinds of observable systems and this overcomes this difficulty we have had in the past.

Miller: I don't find anything wrong with your terminology there.

Newcomb: I would be happy with your proposal. Let's move on to Gerard's question of self-containedness. I would like him to indicate to begin with, what are the presumed criteria of self-containedness.

Gerard: This is what we have discussed before. If you can say what you want to about the system in terms of a system's response to inputs, which inputs are themselves not particularly related to the system, then you have a self-contained system. But if in order to understand and prescribe you have to include things which are not in the description, which are more than unstructured inputs, then it is not a self-contained system and you are dealing with a part of a larger system. Specifically, in your instance, the question would be the extent to which the biological entities involved have to come into your analysis and interpretation of your interaction system.

Miller: Would it be helpful for you to put the diagrams on the board again?

Newcomb (At board): In Figures 1 and 2 A is a person; B is a person with whom he is actually or potentially communicating; X is the object or referent of communication. Solid lines represent actual attitudes. Broken lines represent attitudes as inferred by the other person. And the statements of interdependence among attitudes or orientations which are the units of the system are simply that a change anywhere in one of them is likely to induce changes in any of the others. The events (let's limit this to the phenomenal system at the moment) have to do with the receipt and the transmitting of information. This is somewhat better understood if you take a social system rather than a phenomenal one, in which case, the diagram is represented by Figure 2, where both the actual and the inferred attitudes of A and B toward each other and toward X are shown.

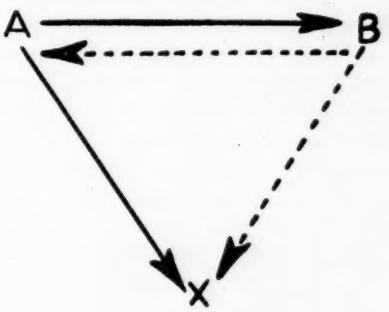


FIG. 1.

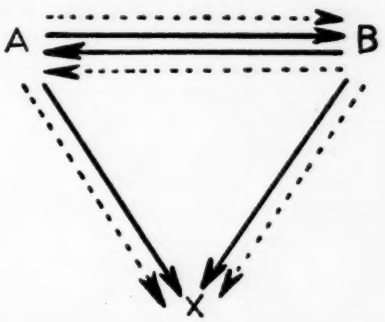


FIG. 2.

It can be empirically demonstrated that under certain states of systems, behavior by which information is transmitted is more likely to occur than under certain other states. And furthermore, that when information is received there are effects upon the states of the system, and these effects can be related to certain existing states of a system.

Miller: Now A is a biological entity; and B is a biological entity; and X is the thing they are looking at. Let's say it's a book. Now what are the lines?

Newcomb: They represent orientations. This is a static description as existing in a

given moment of time. There is no behavior here.

Gerard: Can you deal with the orientations alone as a system without bringing in the individuals A and B?

Newcomb: Without bringing in the individual as a biological organism. . . .

Gerard: As a material entity. . . .

Marquis: And without bringing in the fact that he's an essentially hostile person or anything of that sort, which isn't part of the system as diagrammed.

Easton: By definition, doesn't the term "orientation" refer to a particular aspect of an organism? A has an orientation and so has B, but not X if it is not an organism. So what you are talking about is particular states of biological organisms. Other biological states may not be part of that system.

Newcomb: The orientations are special abstractions of certain aspects of an organismic state.

Gerard: But the total system that you have in order to think about it does include the other. There's no question about the usefulness of abstracting certain things from the total from time to time, but the system of orientations is a partial system and not a whole one. It is not self-contained. Do you agree?

Newcomb: By that definition, I think this is right.

Easton: Well, it's self-contained in this sense: if by looking at this you don't have to go into detail about the other. Let's look at it from this point of view. You have a system of attitudes in regard to X, which is a book. A is trying to sell the book to B, and this is the material basis of your system. You are talking about one part of the total state of the organism that is relevant to X. Now there are other parts of the state of the same organism that are not as directly relevant to X, which, of course, affect the partial state of the organism that is relevant to X. The changes of the other parts of the state of the organism will be inputs to this particular part. For example, if A happened to have some experience just before he came and met B, which gave him further information about B, this would be an input into that part of the state of the organism and it might change his attitude towards B. It seems that you can separate out these specific attitudes of A and B with regard to the book. All the other parts of the state of the organism, as they change, become changing inputs into this particular part.

Gerard: But then I come back to the other question which precipitated all this. Can the subsystem have any dynamics unless you bring in the rest of the system? That is, changes in this are inputs and outputs not into the attitudes or the orientations but into the individuals which leads to changes in attitudes and orientations.

Miller: I wonder if I can summarize what is now becoming clearer in my mind. Newcomb is being a good psychologist. Psychology is an "order-analytic" science where you deal with correlations between observed inputs and outputs at the level of the individual without going down to the level of his subsystems. So, what you call an orientation is a name for a repeatedly observed correlation, or a correlation derived from repeated observations where A acts in a certain way, whenever B and X are in the situation. There is no time involved in these diagrams. That heavy arrow from A to B means that there is at a given moment a correlation between B as an input into A and A's action, whatever sort it is. This is what we call A's orientation toward B. You don't say anything about the causal "dynamics." We could say, if we wanted, that there is a "trace" in A somewhere—we don't know what it is—which guarantees that this correlation will occur. But we don't need to bring this hypothetical trace in. Finally, since we are operationists, we don't say anything about what A thinks to himself about his orientation toward B. All we talk about is what two or more observers could observe in his overt behavior, either his language or his actions. Therefore, the subjectivity of A is not what we are discussing. We are discussing the correlations between a given input, which observers can see, and an output, and we are not considering the relation between that and the possible causal mechanisms.

Gerard: Miller, it seems to me that you are going off on a tangent from the direction

I am trying to go. May I revert to forcing this issue through to a conclusion. We have been over it many times, but I think that we are perhaps ready to say yes or no. Will you allow me to postpone that inner and outer, etc.?

Certainly one part of what you said is relevant, and that is the successive states. And this is part of the dynamism. Can the states of the system be derived from one another without invoking this other part, beyond the subsystem that you are dealing with?

Newcomb: If I understand you, you are saying additional mechanisms are involved here. To this I am inclined to reply: so long as there is regularity of consequences, we don't have to be concerned with the mechanisms. For example, the mechanisms by which spoken sounds are heard, and neural impulses are conducted. . . .

Gerard: Now I didn't raise the question of specific mechanisms at all. It is only that the organism is the mechanism whereby these orientations interact. . . .

Newcomb: The mechanisms by which information affects orientation. . . . Is that what you were calling for?

Gerard: Can orientations affect themselves by any other mechanism in your thinking?

Newcomb: Yes. (explaining Figure 3) Here we have systems of orientations; and here we have a postulated condition of a total system; and here is communicative behavior. The arrows point to possible effects. Receipt of information is considered as communicated behavior. This is one of the things that can change it. But certain combinations—intrasystems combinations—are regarded as stressful or result in strainful changes.

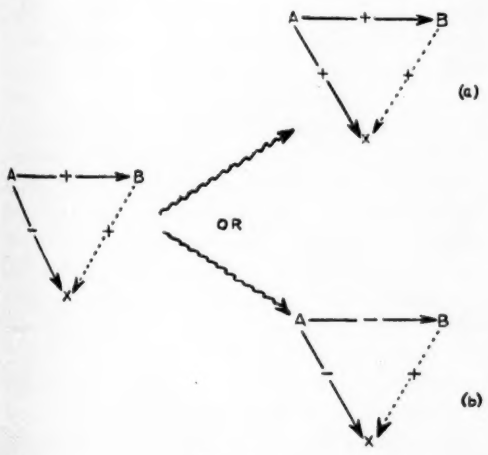


FIG. 3.

Gerard: Doesn't behavior still involve the mechanism of the behaving person? In other words, I don't see how you twist it to avoid introducing the mechanism of the organism as the means of interaction in your system. If you say the system has dynamic properties of successive states, which can lead into one another without the organism being invoked in any way, you are saying this is a sufficient system, and not a subsystem of a larger one in a meaningful way. You haven't shown me that yet.

Newcomb: It's not sufficient in those terms. Of course, the organism is. . . .

Gerard: All right, then you would be willing to accept, at least at this level, that your argument concerns a subsystem of an organism system, and you may be very much more interested in the subsystem than the complete system, but you can't dispense with it. Is this acceptable to you?

Newcomb: Up to your last sentence where you say you can't dispense with it. I think for my purposes, I can.

Gerard: You can't dispense with it and make the system work.

Newcomb: If you want a full and complete explanation of it you cannot. . . .

Gerard: To make it dynamic.

Easton: Suppose you defined orientation as simply the state of the organism, then this would by definition include the organism. So that your problem doesn't arise. Let me give you a concrete example. Suppose X is a rare book, B knows it is a rare book; A does not. A is trying to sell it to B; B is trying to buy it cheap; A is trying to sell at a higher price, but he lacks the information. The attitudes that are represented by the solid lines and the inferred ones represented by the dotted lines are based upon each person's conception of the value of the book. Each organism is presumably in a certain state, and the two organisms are

acting in terms of these states or they are oriented toward each other accordingly. (By oriented, you mean simply that the state of the organism prepares each individual to act in relation to the other individual in a given way.) Now let there be a certain input. A suddenly gets information that X is a rare book. Then you would see the content of the solid lines, and possibly of the dotted lines, change. You would have a change in the states of the organism that are related to the orientation of A to B at that particular moment. But, for the moment, we have excluded all those other states because they don't seem to be particularly relevant. Suppose, however, that someone came up and said, "Now here's a state of the organism that is very relevant and you ignored it." Then I would have to consider this to be another kind of input or broaden our system to include it. And as I see it, what you are arguing, is that all other states of the organism must be included.

Gerard: No. Let me go on and say, what seems to me the logical consequence of this, and then we'll throw it open to discussion. It seems to me, that the way this was coming out was quite clear and decisive. And I thought Newcomb and I were agreeing well enough to formulize this end, and the rest of you could accept this or not as you wanted. But the issue seems, not that between "real" and "unreal," in the sense of material or immaterial, although I suspect that it will always come back to that; nor is it in terms of observable or non-observable, although this is likely also to come back. It is essentially to whether the particular system with which we wish to deal is sufficiently self-contained to be useful when dealt with in that fashion. The main question as to whether this would be useful or not seems to me inherent in whether it can contain dynamics, and so far I have not seen any system which did not contain the material entities, the organisms which are interacting, which sounded as if it were dynamic. If this is true in terms of the orientations or attitudes, it is equally true in terms of the acts. And so an action system is also a partial system which may be very useful to look at from certain points of view, but it will not have any dynamics, per se, and it is equally true for a state system which is simply dividing it in time in another way. So, I find myself again coming back to saying that the adequate system that we have to have in our minds, whether we use all of it, is a system of material entities, of which orientation systems or action systems are subsystems. Whether we can usefully deal with the subsystems will depend on what we are trying to do at any particular time. If we don't need any dynamics, if we are willing to neglect the causal and simply have a descriptive statement, it may be economic and very desirable to eliminate the rest. But to say that this is an adequate, sufficiently self-contained system in the long run. . . .

Newcomb: I would like to hear Rapoport. In the meantime, I am completely opposed to everything you said in the last five minutes.

Easton: I would like to hear from Rapoport too. But first I have something less definitive to say. My remarks were really quite directly relevant to the problem, and as I see it, answered the question that you posed. The system is dynamic because information inputs change the states.

Gerard: Input is into the organisms?

Easton: We define the system in terms of the states of the organisms since by definition we're talking about states of the organisms Not about states in the abstract, but states of the organisms.

Rapoport: It seems to me first that there were two questions involved. First, "Are systems of actions *conceivable* without underlying physical or biological entities?" Second "Are systems of actions *understandable* without the underlying physical or biological entities?" Now, that they are not conceivable, I readily admit. Certainly there must be entities to make the actions. But I think action systems or orientation systems can be *understandable* without the underlying entities. Gerard, do you think that systems of actions cannot be understood without bringing in the carriers of action?

Gerard: No, I said in certain circumstances you may be quite happy to use them.

Rapoport: I can think of some examples. I will take a physical example first. It may

not be directly relevant, but it is analogous. Recall the change of thinking in electrodynamics that occurred in the middle of the nineteenth century from considering the interaction of charges that is, "entities" to considering the interactions of fields. Now fields are not conceivable without charges. There must be charges to give fields. But the system turned out to be more understanding if one confined oneself to the interactions of fields. Charges then become "singularities," that is, discontinuities in the fields. Fields may be taken to be analogous to actions, and charges to individuals. Electromagnetic systems, it turned out, could be more profitably studied in terms of "action" interaction than in terms of interactions of the carriers of action.

The other example, which is perhaps more relevant to our purposes, is the game of chess. I maintain that a game of chess can be analyzed dynamically, that is in terms of successive moves, without any reference to the players. Of course, you may understand more if you understand the psychology of individuals, but a theory of chess is much easier to construct in terms of the succession of moves, without reference to the individuals. The moves themselves have a certain logic. And I wonder if certain action systems cannot also be regarded as self contained without reference to the actors. In other words, quite apart from the biological entities which give rise to them, perhaps action systems or orientation systems have a logic of their own. If they do, it is important to look at these "logics" because of tremendous saving of effort that one is able to effect.

Gerard: I completely agree.

Newcomb: I was going to say something along the same lines. I think we might have a semantic trouble here over the word "dynamic" which to you obviously means biological. It doesn't mean that to me.

Gerard: No, but I do think it means "material"...

Newcomb: But it doesn't to me. I use the term "dynamic," I think, consistently and clearly, simply to mean a statement of the conditions under which change and stability occur. That's all I mean by dynamics.

Gerard: However, change is of *something* in time.

Newcomb: How can "nothing" change? If you have nothing, it won't change.

Gerard: That's quite right. It is the interaction of the changing elements which demand some material substratum in my way of thinking.

Newcomb: To which I would say, in order to get understanding, as Rapoport says, it's often perfectly reasonable to short-cut, to take for granted, the intervening mechanisms. All psychologists do this all the time. By your definition, I think there is no such thing as self-contained psychological system, and you may be right.

Gerard: Now I did include the statement that there may be instances in which it is quite useful and parsimoniously desirable not to go outside of your system. But it remains a partial system.

Newcomb: In a sense, all systems are partial.

Gerard: I suppose every system is, but there are still very real differences in the degree. It's like that history-geology story.

Marquis: Well, then, the criterion is usefulness—the degree of understanding?

Gerard: I think so.

Marquis: But we could call it a system, couldn't we?

Gerard: Again we are not concerned about the words. The question is, under what conditions is it worthy of our attention? This is the only question that I am asking.

Marquis: But you are saying more. You're saying an orientation system has to be thought of as a subsystem because we hadn't gotten down to any material entities interacting.

Gerard: I am saying that in effect, and this is obviously where I can be way off, from the psychologist's point of view. But I still think it is useful to think of an orientation system as a subsystem rather than a system. I can certainly think of the individual man as a system, although I am well aware that he is also a subsystem. Now, if I can think of attitudes interacting without bringing the man in, I'll be very happy. So far, I can't do this.

Newcomb: I would then summarize your

view that a system of orientations is a less complete system, than that composed of entire organisms. I don't think there is any doubt about that. But partiality is itself a matter of degree.

Gerard: Are we all in agreement? I think the issue is pretty clear. It seems to me also that for once we have had a pretty good argument and clear agreement, although there may be differences in weightings and emphasis.

Marquis: No, I still have one question. Can you have an action system or do you have to have a system with entities in it? It seems that A and B represented persons, and I want to ask whether they are the entities of the system? I don't think they are.

Newcomb: No, they are foci or sources of action.

Marquis: They serve to define the arrows representing attitudes, attractions, etc. Now as the arrows change. . . .

Gerard: They don't change except by acts. . . .

Marquis: This becomes the point. Why don't they? Suppose I assert that they do, that there are inputs to *attitudes*, not to people, but to attitudes. When I can correlate any event with a change of attitude, I can talk about an input to the attitude, and all I measure is the attitude. I don't measure the person. And all I can conclude from this system is that if this input event changes one attitude, then another will change concomitantly.

Miller: Let me suggest that perhaps there is an answer to that one. Psychology differs from physiology in that it is "order-analytic," whereas physiology in relation to behavior is "causal-analytic." Psychology has also traditionally differed from physiology in that psychology has used subjective terms like orientation. We recognize now that we only get at the meaning of such terms by observations, operations, and conceptual agreements, but the terms we use are nevertheless subjective terms. Now there isn't any reason that I can see why Newcomb should not substitute the word "trace" for the word "orientation." If he really believes, which I know he does, that the objective view is simply a different epistemological route to the same phenomena, then he might advance behavioral science by giving up the whole set of subjectivist terms and use only objectivist ones.

Gerard: Does this fit your point, Marquis?

Marquis: I may be a sucker for terminology, but I didn't think it would change my point of view if I just changed the wording of it.

Gerard: Can we make sure we have your point. . . .

Marquis: Can you imagine a system in biology in which the entities, the things that change with inputs and interact with each other, were things like ratios, proportions and excretion rates, and specific gravities. Now, those are not entities, are they?

Gerard: I can answer this quite explicitly because this kind of thing has come up over and over again, and in every case, it has proved unsatisfactory and has been superseded by an actual material picture. One of the best examples has been in the equations in the excitations of nerves, in which Lapique, among others, tried to get excitation in terms of a difference in concentration. It didn't work. You have to have some actual amounts. That's the only example that pops into my mind, but I think this is a fair generalization. (To Rapoport) Will you agree that such has been the case in physical science or not?

Rapoport: In physical science, I am afraid I must agree. But let me qualify it. It is the case because the final step, that is the connection to concrete physical entities, can in most instances be made. And if it cannot be made, one expects that in time it will be made. In situations where it is almost hopeless to make connections, I would be opposed to using physicalist props. The reason I would not like to use "trace" for "orientation" is because we cannot, at this time, describe hypothetical traces, but we can describe orientations, at least to a degree.

But even in certain branches of physical science, when you try to come to grips with material entities below a certain level, you just lose sight of them . . . they evaporate, leaving you only with probability waves

with imaginery amplitudes and other eerie abstractions.

Gerard: This was in response to your particular point, carry on. . . .

Marquis: Well, that essentially is what that is. You see, an attitude is like a ratio . . . a relationship. . . .

Gerard: I would say in psychology this is a satisfactory arrangement.

Marquis: I wouldn't say satisfactory. What I am really asking for is the kind of guidance which you offer by saying what has worked in other fields. It seems to me that there are two ways that one can formulate what Newcomb is talking about. One is that there are individuals A and B and the thing X. And that we are going to talk about only one measured aspect of A; that it is a reduced system in that sense. That keeps it an entity system. . . .

Gerard: I think you introduced a word there I have been searching for. We can speak of whole systems, part systems, or reduced systems. I think reduced system may be a useful term.

Marquis: Or the other way is to take the attitudes themselves as the entities, the interacting things of the system and not talk about individuals. Now, I don't know whether this makes any difference.

Gerard: It is going to come down completely to the kinds of things we can do when we try to think of them.

Marquis: You mean the only answer is pragmatic?

Gerard: I think so. If Newcomb can give a usable picture of what he is talking about without invoking the properties of men, other than the orientations or any other ones he explicitly enumerates. . . .

Ruff: You just said, several moments ago, that the group had reached agreement, and I wasn't sure on what agreement had been reached.

Marquis: To summarize—The components of the system do not have to be real in the naive sense, they may be inferred from observation or may be relationships based on observation. Newcomb's system could be formulated either with entities such as individuals or with attitudes and orientations, and the dynamics of the system in the latter case, would deal with inputs affecting attitudes and concomitant changes in them. Which of these two formulations is best, cannot be categorically stated. It is a question of pragmatic usefulness in the development of science.

Easton: Just to tie this in with our earlier discussions of other kinds of systems, for example, systems of ideas. Systems of ideas are also effects (in the same way that systems of attitudes are effects) which can be, for the time being, examined independently, until the time when we have techniques for studying the underlying system of entities more thoroughly.

Marquis: One criterion of a system is that it is self-contained; this means that the interaction of the parts of the system and its changes of state can be understood in terms of stipulated inputs and outputs. Either of the possible formulations could possibly produce a self-contained system, although it remains to be found out in what degree.

Gerard: The degree to which it's useful to treat it as a complete system of self-containment. This is one time that we have gone over this and that I feel happy about having done this. Thank you for raising the questions as you did. Will you go on to the next question?

Newcomb: Let us return to the diagram (Fig. 3). Assume that A's attitude toward X is "positive" (denote by a plus sign on the solid line). B's attitude inferred by A, on the other hand is "negative" (denoted by a minus sign on B's dotted line toward X). At some time the discrepancy is discovered. This discrepancy defines a state of strain. The assumption is that in such systems there are constant forces operating to reduce the strain.

Gerard: May I interrupt because I am not clear there? This is the inferred attitude of B on X. A likes B. A likes a certain girl. A thinks that B doesn't like her. This seems to me much less of a strain than if A thinks B also likes her.

Newcomb: A competitive situation can also be handled by this method, but it's more complicated. Let's take something not competitive. Now there are two conceivable mechanisms by which the existing

state of strain can be reduced. A very common one is by the induction of behavior. Communication is a common solution. Here, either A persuades B to his point of view or is, himself, persuaded to B's point of view. And in one case or the other, discrepancy is reduced. But this doesn't always happen. It may be that communication only makes things worse. In that case the change may come about in A's attitude toward B, changing the plus in the arrow from A to B to a minus. There are, thus, alternative methods of strain reduction: Change through communication, and autistic change, resulting from the system's own tendency to reduce strain. In this case perhaps we can speak of inputs from one attitude to another. I don't know whether this makes any sense.

Rapoport: Why not? A's inferences about B's attitudes certainly may change A's attitudes toward B.

Newcomb: And you would describe this in a language of input and output—from one orientation to another?

Gerard: You're bringing A in, not the orientation.

Newcomb: These are all orientations of A. And you're saying that one of A's orientations has input for another of A's orientations? I wouldn't want to . . . this doesn't sound right to me.

Easton: What you would have there are "withinputs" from one subsystem to another subsystem if you want to analyze it in those terms.

Miller: But then you would have to have A again in order to have a subsystem.

Marquis: You have to have an input to create a strain but the resolution of the strain doesn't require an input, does it?

Rapoport: Is this a strain situation now? (Referring to B in Fig. 3)

Newcomb: No, this is now at rest.

Rapoport: How do you tell from a diagram of this sort whether there is strain or not?

Newcomb: The general principle is that if $A \rightarrow X$ and $B \dashrightarrow X$ are discrepant in sign, then a positive state of $A \rightarrow B$ is strainful and a negative state is not.

Rapoport: I see, so it's the algebraic product of the first two that determines the last. Suppost that the first two were both minus?

Newcomb: It doesn't matter what the sign is. It's the likeness or unlikeness.

Gerard: But as you have formulated it, you still have not excluded the individuals.

Newcomb: They are always there as a substratum. I want to raise two questions: First, are there other examples of systems in which steady states can be achieved either by intrasystem exchange, interaction, or by input or output across boundaries?

Miller: There are millions of examples.

Newcomb: In biology or physics?

Rapoport: Take the distribution of concentration of some substance, in some bounded region. If it is a closed system, then a certain equilibrium concentration will be achieved throughout the region. But you can also achieve a steady state by having a certain input and output, through the system, and then you can have a distribution of concentration which is not an "equilibrium" but which is nevertheless a steady state.

Newcomb: So there's nothing unique in that respect?

Miller: I wish you would say something about how this system changes in time.

Newcomb: Let me draw a picture of how I handle the business over time. (Fig. 4) Here is a system state having certain properties as a result of which, starting at a given time, there will be communicative behavior —and under certain conditions the state of the system will change over time. At any given time, theoretically, communicative behavior is an outcome of the just preceding system states—and system states at any preceding time are outcomes of just preceding communication and/or autistic changes which are also indirect consequences of communication.

Miller: Then what you are saying is that there can be changes in the system state (*a*) due to the reduction of strains (*b*) due to forgetting, (*c*) distortion, etc. In addition,

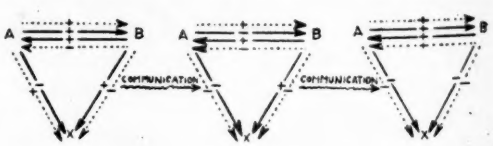


FIG. 4.

there can be changes in the system state due to new information coming in.

Newcomb: And the new information coming in is the consequence of things you do in general to get it. I would like to know how this relates to processes occurring in other kinds of systems.

Gerard: As you present this, this is a beautiful example of why, what you are talking about is reduced or a partial system because you do keep bringing in these other things. Now if you do—would it be useful to you to formulate it so as not to bring them in—and would it work for you? If you would like to try this I would love to see it done.

Miller: Not bring what in?

Gerard: The acts of the organisms—the behavior and the communication.

Newcomb: I think I would rather pay the price of having only a partial system than the price of bringing everything in.

Gerard (to Marquis): Do you feel that way too?

Marquis: No, I don't like it. Because he can add anything he wants to. I would like to have him define the scope of the system ... tell us the boundaries ... and see how much you can explain with that. I like the second kind of formulation in which the attitudes and not the individuals are the entities.

Rapoport: If A and B agree, the chances are that one likes the other. And if they disagree, they either come to an agreement or one does not like the other. Now what do we do with this set of conclusions? What is the next step?

Newcomb: I feel forced to answer in pretty general terms. One is presumably interested in properties of an individual or of a group. And this proposes that among the relevant properties of an individual to study are his orientations. But these occur not in isolation but as parts of systems. If one is interested at the level of molar behavior of individuals, that is, distinct from molecular levels, then, I think, the most useful ways of doing this is in terms of systems of orientations and attitudes as variables. The task is to list a set of predictions as to what kind of behavior will occur or how attitudes will be changed.

Marquis: What you said demolishes my position. You want to expand and extend this formulation to explain the behavior of individuals and individuals in interaction. And you therefore abandon the idea of a tight little system of particular orientations that you started with. You want to add more variables, more aspects of individual behavior.

Gerard: This is in response to Rapoport's pointed question. What have you done with this that is more than simply a qualitative statement which can be made directly from observation.

Easton: I think Newcomb is selling himself short. Because he has done a great deal more.

Gerard: I think this would be a good time to indicate some of the predictions. ...

Marquis: What about distorted perception? Is it included?

Newcomb: It is. I have been going over some dirty data which indicate beautifully that within one group of 17 men where I've made some hundreds of thousands of observations ... that the probability by which judgments of other people will be distorted is a function both of frequency of communication with them ... and positive attractions toward them. Then, of course, it turns out that these two in turn are interrelated and you get into a nasty problem of separating these out. There is a lot of work by Festinger which shows that the likelihood of persuasive communication varies directly as the positiveness of interpersonal attitudes.

Gerard: You are beginning to give these numerical values. Do you make any unexpected predictions?

Marquis: I am coming gradually and reluctantly to a position which might represent a reconciliation: that it is more in the interests of psychologists to look at individuals as systems, provided we can deal with incomplete or reduced or part systems, than it is to build systems of actions or attitudes and so on. And this would permit us to stick to the hierarchy of material systems; organs, individuals, groups, etc. You see, I would be terribly embarrassed if I tried to get a conceptual system of personality or motivation into this hierarchy. What I am now saying is that it sounds acceptable to me to consider conceptual

systems as reduced systems; not subsystems but incomplete systems. And the reason for this, is that I always want to move, to the full system for the individual. I would not want to stay with those part systems. Newcomb, too, reveals so clearly that he isn't really interested in these attitudes per se but that he's interested in them in relation to other attributes, aspects, and characteristics of people; in order to explain people. Our system will always be about people and material things but we can talk about reduced and incomplete aspects of them.

Gerard: So you are saying you want the privilege to hold the temperature of the gas constant and vary the pressure to see what happens to the volume, and you want the privilege of holding the pressure constant varying the temperature to see what happens to the volume, but you recognize that sooner or later you will put them all together.

Rapoport: I disagree, I don't think that's what he wants.

Gerard: Well, I will withdraw what I said.

Marquis: Do I want what I want?

Horowitz: The level of your part system is defined by what it's part of.

Marquis: I meant to say that.

Easton: The trouble is that we have been playing tic-tac-toe on a two-dimensional plane. I like to play tic-tac-toe on a three-dimensional plane.

Rapoport (to Marquis): Isn't what you want the privilege to slice the world as you see fit at the moment?

Miller: And also to know at what level to slice it?

Marquis: I have another question then. Is there any difference between Newcomb's system and an organ system like respiration? Would they be at the same level?

Newcomb: This is what I want to go on to.

Miller: May I try to answer that? Yes, there is a difference.

Gerard: There is no difference.

Newcomb: It's a second order of difference.

Gerard: You are trying to define the difference between a subsystem and a reduced system.

Miller: That's right. A subsystem of a real system is a special class of a reduced system. But I think that there are many reduced systems which are not subsystems in this sense.

Marquis: What would the temperature-regulating system be?

Miller: The temperature-regulating system is a subsystem.

Marquis: But an animal sweats all over.

Miller: Yes, but the regulating system for temperature in a house is the thermostat and the furnace, although the whole house has a temperature.

Marquis: Wait a minute . . . is the output of the regulating system a nerve impulse? Or is it sweat? I think the boundaries of your system have to include sweating, the intake of water, the movement of limbs. I don't see how that's a subsystem. I don't see how it's any different from "personality."

Newcomb: I am going to argue that all systems we use are reduced systems. Nobody ever could deal with a complete system.

Gerard: Let me remind you of one thing that you did agree on. We can have subsystems which are such in terms of their origin, for example, the different tissues; subsystems, in terms of their structure, such as the organs; and subsystems in terms of their function; such as the temperature-regulating system.

Marquis: What would Newcomb's be?

Gerard: Is it function or structure of a mixture?

Newcomb: Function, I'd say.

Marquis: And like temperature regulation, it involves structures but it's defined in terms of function.

Ruff: I don't understand the reduced system. When you first started using it, it tied up with the issue that has just been discussed, namely that of abstracting from a material system. And now it seems to be tied in with subsystem. I'm a little confused.

Gerard: A reduced system is one aspect of a system without its having been defined too closely.

Marquis: With some of the variables but not all of them.

Miller: If we have a glossary I think it would be desirable to include a definition of

"reduced system" and a definition of "subsystem."

Marquis: Incidentallly that term probably isn't very good because reduced implies incomplete.

Newcomb: "Abstracted" is most accurate of all these words.

Marquis: Except that even the complete one would be abstracted.

Rapoport: The system organized in accordance with criteria different from those we have been using. Abstracted in a different way. All systems are abstracted; some are abstracted so often that they are imagined to be concrete. Systems can be abstracted in various ways. When the abstraction is unusual to our way of thinking, we have trouble classifying it, and since this kind of abstraction doesn't fit into previously made abstractions, trouble arises. Let us recognize once for all that we abstract in different ways, and the reason for making a particular abstraction is the usefulness of that abstraction.

Gerard: You are rephrasing the old statement that no matter how you slice it, it is still baloney.

Rapoport: Exactly.

Newcomb: I would be interested in knowing whether the kinds of problems I meet here are typical of problems other scientists face. Also I do like to assume that we will derive no knowledge about systems except from studying specific systems of many kinds, and this is one.

There is a two-level problem here. Just as I tend to view an individual as a system of orientations, so I tend to view a group as a system of systems of orientations. I can and do make predictions at both levels, but what I am not sure of is whether I am really saying the same thing in both cases using slightly different language. Perhaps I had better take a specific illustration of a two-level problem, because this is where I really want help. At the level of the individual, we can say that, given perceived discrepancy with regard to an orientation of importance, that is, given a strain, the likelihood of communication is very great. Now let's take a group of any size. Here you would simply say, without regard to single persons, but with regard to relationships among persons that the likelihood of communication with regard to a given subject within this group will vary directly with discrepancies in attitude and attraction. Take a two-person group just to make it easy. Again there is strain to the extent that there is discrepancy about a relevant object. I haven't said much about relevance, but it's absolutely essential to my meaning that there is a joint dependence on the part of both upon this object. A condition of maximal strain is illustrated by two men on a life raft in the middle of the ocean. The wind is coming up, they disagree completely on how to manage the rudder. The use of the rudder is obviously relevant because they are jointly depending upon it. There is attraction between the two men but it stems not from liking in this case, but from the need of each person for the other. The raft can't be managed by just one person, so they have strong attraction, a highly relevant object, the rudder, and if they disagree completely, a maximal strain results.

Gerard: Real discrepancy or perceived discrepancy.

Newcomb: Real discrepancy. Under these conditions you predict that there will be communication about the rudder. And you don't care who starts the communication. If you put it in terms of a social system, you simply say that, if there is strain anywhere in the system or if there is perceived discrepancy anywhere in the system, then somewhere in that system communication will occur. And the likelihood is that this communication will have consequences that either the discrepancy is removed or something else will change. Now this is a statement which can be made without reference to any particular individuals in it. You merely say it will happen somewhere in the system.

Gerard: Whereas if X is not vital it may be the other way—they will stop communicating as a result of the strain.

Newcomb: A lot of strain is tolerated. The likelihood of communication varies with the importance and the relevance of the discrepancy.

Rapoport: May it not happen that the

importance of the interpersonal attitudes is so great as to outweigh the effects of a strain in the system?

Miller: I think that's a useful diagram for predicting outcomes, especially if you quantify the strength of the attitudes. But here we get to the empirical problem—quantifying the strength of these solid and dotted arrows in terms of comparable units. If you could only do that you would really have a testable hypothesis.

Newcomb: This is a major methodological problem.

Gerard: Miller, you said this is a fine diagram for predicting something. Will you predict from that diagram that they will or will not speak to each other?

Miller: It would depend on the numerical values, as I was saying.

Newcomb: In the case of a shipwreck you surely would predict that there would be communication if you knew this. The point I wanted to raise here is that I have attempted to make statements about a social system without having introduced any new data pertaining to a specific social system. Is this legitimate? In the one case I am making a prediction about a person saying that this person will communicate. In the other case, I merely say there will be communication somewhere in the system. And this is done without any additional information or additional observations.

Rapoport: It's done all the time. When a strain is discovered, an engineer can often predict that something will give without being able to specify what.

Gerard: I was just going to say that what you have really done is thrown away information. In your simple system, you said that A will do something; in the other you say A or B or both will do something.

Newcomb: Does this often or always happen when you move up one level?

Rapoport: I would think that's the usual sort of thing.

Gerard: I am not sure it's by any means necessary though. You may be able to go farther and specify the conditions under which A or B will make the first move.

Miller: May I ask two questions about Figure 2? The dotted line from A to X means what A thinks B thinks about X?

Newcomb: This line means A's orientation as B thinks it is.

Miller: Well, if that's what that means, it has to do with our problem in two ways. In one case, we put two subjects in a separate room. They are talking about similar phenomena and they have similar orientations; they also are active but neither knows that the other exists.

Newcomb: Neither knows that the other is there.

Miller: In the other case, each is aware of the other's existence and takes into account not only his own attitudes but also the inferred attitudes of the other.

Newcomb: Then you have created a social system. In outlining the questions I would like to raise, I included one at the end as to whether there are any emergent properties of a social system as compared with an individual system. There is at least one, there is the property of objective agreement or discrepancy which doesn't exist in the other case. Are there others? Is it invariably true—I should think it would be—that if one has a system made up of subsystems then surely the more inclusive system has some properties, that the less inclusive system does not?

There is a serious side to this unabsorbed gorge of science. It has given our people a bad case of indigestion. It lies in the public stomach and troubles their dreams. They do not know enough to know good science from bad. . . . What the world needs is not less science, but more knowledge of what science is and what it can and cannot do.

—E. P. Lyon in *Sigma Xi Quarterly*, December, 1936

ABSTRACTS OF CURRENT LITERATURE WITH INTERDISCIPLINARY IMPLICATIONS

Prepared by Members of the Staff of the Mental Health Research Institute

Adam, A. O. **Crossbar tandem as a long distance switching system.** *Bell System Tech. J.*, 1956, 35, 91–108.

Crossbar tandem switching is coming into general use in the Bell System. This paper discusses the aspects of this method which apply to nation-wide dialing, the establishment of a hierarchy of sectional, regional, and national centers for telephone communication, the selection of alternate routing to bypass busy lines, code conversion from one system to another, and adding and deleting digits to the call code as required. A survey paper.

See also: *Truitt, C. J.* Traffic engineering techniques for determining trunk requirements in alternate routing trunk networks. *Bell System Tech. J.*, 1954, 33, 277–302. (*C.F.*)

Carroll, John B. (Ed.) **Language, thought, and reality. Selected writings of Benjamin Lee Whorf.** Cambridge, Mass.: The Technology Press, & New York: John Wiley & Sons, 1956.

Five of the papers in this volume have been reprinted before. They are "Language and Logic," "Science and Linguistics," "Linguistics as an Exact Science," "The Relation of Habitual Thought and Behavior to Language," and "An American Indian Model of the Universe." Among the papers reprinted here for the first time are "Language, Mind, and Reality," "Gestalt Technique of Stem Composition in Shawnee," and "The Punctual and Segmentative Aspects of Verbs in Hopi." These eight papers embody Whorf's view that the 'Weltanschauung,' 'metaphysics,' or perception of the world of a group is determined by the language that it speaks.

"A Linguistic Consideration of Thinking in Primitive Communities" is a general statement of the importance of linguistics for psychology and anthropology. This is the paper that introduced Whorf's concept of "cryptotypes." "Discussion of Hopi Linguistics" expands on this idea.

"Language: Plan and Conception of Arrangement" presents Whorf's schema of analytic categories to be applied to any language. "Grammatical Categories" and "Some Verbal Categories of Hopi" respectively supplement and illustrate this analytic schema.

The remaining five papers in the volume include two on psychological topics, "On Psychology" and "On the Connection of Ideas"; two on the decipherment of Mayan hieroglyphs, and one "On Linguistic Factors in the Terminology of Hopi Architecture."

The volume has a foreword by Stuart Chase and an introduction by the Editor. This latter includes a brief biography, an analysis of the development of Whorf's thinking, and a statement on the present state of work on the linguistic relativity hypothesis.

There are three bibliographies: Whorf's published works, a selection of his unpublished manuscripts, and selected books and articles relating to Whorf. (*A.H.*)

Danziger, Lewis, and Elmergreen, George L. **The thyroid-pituitary homeostatic mechanism.** *Bull. math. Biophysics*, 1956, 18, 1–13.

In this paper equations are set up to describe the production of thyroid hormone and its homeostatic regulation by feedback action from the hypophysis. More precisely, it is assumed that the pituitary gland produces thyrotropin which activates an enzyme of the thyroid gland, and that the rate of production of thyroid hormone is proportional to the concentration of that enzyme.

In the following, let

θ be the concentration of thyroid hormone,

π be the concentration of thyrotropin,

E be the concentration of activated enzyme,

and all other quantities be positive constants. Then the equations are derived as follows:

$$(a) \qquad \frac{dE}{dt} = m\pi - kE,$$

that is: the actual rate of production of the enzyme is proportional to the concentration of thyrotropin, in addition we have the "decay term," $-kE$.

Similarly:

$$(b) \qquad \frac{d\theta}{dt} = aE - b\theta$$

which means that, except for the decay term $-b\theta$, the production rate of thyroid hormone depends linearly on the concentration E of activated enzyme. (This is assumed to hold unless severe iodine deficiency limits the rate of iodine utilization and, therefore, the production of thyroid hormone). As we have seen, E is determined by π, but π itself is, in turn, assumed to depend on θ. The authors postulate that a positive contribution to $d\pi/dt$ will occur only if θ is sufficiently small:

$$(c) \qquad \frac{d\pi}{dt} = -g\pi + c - h\theta \qquad \text{for } \theta \leqslant \frac{c}{h}$$

otherwise:

$$(d) \qquad \frac{d\pi}{dt} = -g\pi \qquad \text{for } \theta \geqslant \frac{c}{h}$$

We may say, from Equations c and d, that large amounts of thyroid hormone decrease the production of thyrotropin, until, however, a critical concentration, $\theta^* = \frac{c}{h}$, is reached, above which $d\pi/dt$ is independent of θ.

The paper deals mainly with the study of the system formed by the four Equations a to d above. More particularly, attention is focused on the variation of θ, the concentration of thyroid hormone, with time. Although each of the equations is linear within its range of validity, the system itself cannot be considered as a linear system of differential equations because Equations c and d together are really equivalent to a nonlinear equation for $d\pi/dt$. This has the important consequence that in certain cases the solution $\theta(t)$ may be of oscillatory nature.

Under normal conditions the hormone and enzyme levels will reach a steady state which is given by the equations

$$(e) \qquad \pi_s = \frac{h}{g}\frac{C}{(1+K)} = \frac{k}{m}E_s$$

$$(f) \qquad \theta_s = \frac{CK}{(1+K)}$$

where π_s, E_s, and θ_s denote the steady state values; $C = c/h$, and $K = ahm/bgk$.

The authors discuss the implications of these last two equations. In their opinion toxic goiter is characterized by a higher than normal value of C (which leads to abnormally high levels of thyroid hormone *and* of thyrotropin). On the other hand, malfunction of the production mechanism in the pituitary gland, and hence low values of C, causes the pituitary form of hypothyroidism, associated with low values of thyrotropin.

Further analysis, however, reveals that the steady states are reached only if the constant K, introduced above, is smaller than a certain critical value K_c. If $K > K_c$ the hormone levels will show sustained oscillations. Such oscillations in the basal metabolic rate (thought to be directly correlated with periodic variations in the concentration of thyroid hormone) are actually observed in the mental disease known as periodic relapsing catatonia.

It is of course not possible to change the value of K; the authors show, however, that the administration of sufficient thyroid hormone will abolish the periodic variations in its concentration and produce only little change in the steady state value θ_s. It should, therefore, be an adequate treatment for periodic relapsing catatonia.

If thyroid extract is administered in daily doses such that the effective rate of assimilation is a constant, R, the system of equations previously set up must be modified to account for this new "exogenous" source of

thyroid hormone. For example, Equation b becomes

$$\frac{d\theta}{dt} = -b\theta + aE + R.$$

It can be shown that the modified system is perfectly stable (no oscillations) provided only that $R > bC$, i.e. if the administered dose of thyroid extract is sufficiently large. A steady state is ultimately reached in which the production of thyrotropin is completely inhibited and the thyroid hormone is supplied entirely by the external source.

From their equations the authors also indicate how to avoid dangerously high concentrations of thyroid hormone (which may be brought about by the external dose together with the transient endogenous concentration before the steady state is reached).

Finally, the authors announce that they have developed a method (to be published) allowing them to perform the needed computations quickly and to estimate the unknown constants which appear in their equations. (*E.T.*)

Hainer, R. M., Emslie, A. G., and Jacobson, Ada. **An information theory of olfaction.** *Ann. N. Y. Acad. Sci.*, 1954, 58, 158–174.

It was found that between the olfactory lobe (or bulb) and the cerebrum there are 24 main paths of nervous fibers. Disregarding, at first, the fact that each of the 24 "main lines" is itself composed of many fibers, the authors consider that by the "all or none" law of nervous impulse transmission, each of them can exist in two different states: "on" (i.e. excited) or "off" (nonexcited). It is further assumed that any given odor corresponds to a specific combination of "on" and "off" in the 24 main lines. There are, then, 2^{24} (about 16×10^6) possible combinations or "messages" that can be transmitted. The number of different odors is estimated to be $\geq 10^4$. However, the communication system may not work at its full capacity; moreover, the incoming pulses are random, whereas, in order to distinguish clearly between all the 2^{24} combinations, the "on" and "off" patterns should be synchronized. It is assumed by the authors that the so-called granular cells perform this task of synchronization within the communication network. This assumption has not been verified experimentally, but certain known facts indicate that it might be true.

A feedback mechanism in the nervous pathway accounts for odor identification right at the threshold of detection.

The larger the concentration of odoriferous molecules impinging on the receptors, the larger will be the number of excited fibers in those of the 24 main pathways which are "on". Thus the two aspects of discrimination (quality and quantity) are treated separately in this model. In the neural pathways of olfactory reception there is a sort of "bottleneck" (the situation is similar for the optical pathways): the number of nerve fibers in each of the 24 main lines is about 2.1×10^6 between the external receptors and the olfactory lobe, but only about 1,900 between lobe and cerebrum. The authors assume that the connections within the olfactory bulb are such as to satisfy the following two laws: (*a*) the Gaussian law of fluctuations and (*b*) the Weber-Fechner law.

From these assumptions they then derive the following conclusions, which represent the predictive value of the theory.

1. There are 30 perceptual levels of intensity, or 29 j.n.d.'s.

2. A 52% increase in concentration is required for a j.n.d.

3. The saturation concentration is 190,000 times threshold concentration.

The authors give one set of experimental results to show that their conclusions are verified. Adaptation and fatigue are also discussed in terms of this model. (*E.T.*)

Lee, C. Y. **Analysis of switching networks.** *Bell System Tech. J.*, 1955, 34, 1287–1315.

Using a simplified model, an analysis of switching networks is presented. Methods for finding characteristics of a network such as blocking probability, retrial and connection-time distributions are given. The problem of equivalent crosspoint minimization is also considered. Ten references to other

papers in the field are included. The treatment is primarily mathematical and requires an understanding of queue theory, probability, and some background in telephone circuitry. (*C.F.*)

MacKay, Donald M. **Towards an information flow model of human behaviour.** *Brit. J. Psychol.*, 1956, 67, 30–43.

The designs of certain self-organizing systems are discussed with a view of tapping a source of conceptualizations for a theory of the human information-handling systems.

The author begins with a system incorporating the minimal requirements for goal-directed activity: a receptor, sensitive to the discrepancy between the output variable and the goal, which sends a "mismatch signal" to a control unit, which adjusts an effector unit to give a response, which brings the output nearer the goal.

The inadequacy of this scheme is noted for representing goal-directed activity which embodies more than "logically deductive operations in a fixed calculus," in other words, systems which, like human beings, "invent new categories and novel hypotheses, to meet an unforeseen flux of events." Accordingly, the author proceeds to design self-regulating systems which, as in Ashby's Homeostat, are guided by a trial and error process. The problem lies in the proper organization of the trials. An automaton, which runs along the road where it encounters only right and left 90° turns at random, can operate on the principle of selecting right and left turns at random, correcting itself each time it is "wrong." But if for long stretches, sequences RLRL . . . frequently appear, efficiency is gained if each selection is the opposite of the preceding correct one. If, however, the sequences themselves alternate (e.g., RLRL . . . and RLLRLLRLL . . .), a second order control is required, which selects among the available *patterns* of response.

However, in all these examples, it is the designer who foresees all the possible situations in which the organism may find itself. For a greater resemblance to intelligence, the automaton should be able to "discover for itself the statistically stationary features of its environment, and develop its own characteristic organizing sequences as a result of its experience, starting if necessary from scratch."

The author gives an example of a "statistically self-organizing system," which simulates such ability. The central feature is a probabilistic "bias" in the elements of the system instead of predetermined rules of response, so that the response at any time is essentially a matter of chance but with the probabilities of the various responses changing as a result of "experience," certain ones being increased and others decreased. Such a system can not only "learn" the "correct" response over a period of trials, but can also "unlearn" the response as the situation changes.

The extension of these statistical principles to pattern learning is straightforward but, of course, necessitates more and more involved design as the range of possible situations increases.

In addition to the reinforcement of correct responses by their "success," another principle of organization is proposed, whereby the reinforcement also depends on the frequency of the response. Such an arrangement will favor "routine activities," i.e., those involving closed cycles. Closed cycles nearest to the correct patterns will be favored and the system will learn not necessarily the exactly correct patterns but rather patterns composed of short closed cycles, as approximations to the correct patterns. It will exhibit a primitive "gestalt perception."

Finally, a third principle of organization is advanced, namely, the "antimonopoly" principle, which counteracts the "freezing" of successful patterns of behavior (essentially by a "forgetting" device), and thus leaves room for flexibility and reorganization.

Having described the general principles which must govern the design of self-organizing systems, the author raises the question of the time scale necessary for successful adjustment to occur in a complex situation. Obviously, if the degrees of freedom of a system are many, the system is capable of fine adjustments, which, however, take correspondingly great stretches of

time to "take hold." Accordingly, it is proposed that the system start out with relatively few degrees of freedom and thus with the capacity of making crude but rapid adjustments, the degree of freedom being then generally increased in number, leading to greater refinement and conditionality of response. It is noted that this principle bears a realistic comparison to the human situation.

The paper ends with a speculation concerning the counterparts of symbolization, "thinking," and "imagination" in automata of this type. In particular, it is pointed out that allowing the automaton to "run free," i.e., without evoking the corresponding effector activity, may make possible the continuation of the self-organizing activity, which might then be taken as an analogue to the purely mental activities of higher organisms.

The question of functional disorder as an analogue to human mental disorders is raised, but its discussion is deferred. (*A.R.*)

Miller, George A. **The magical number seven, plus and minus two: some limits on our capacity for processing information.** *Psychol. Rev.*, 1956, 63, 81–97.

Viewing the human observer as a communication system, the author discusses the limitations of this system in the transmission of information and as some of the mechanisms which counteract these limitations.

Experiments on absolute judgments have shown that human observers can accurately discriminate between four and ten classes of stimuli varying along a single dimension, with a modal value in the neighborhood of seven. In other words, the "channel capacity" of the human observer is limited to the transmission of around 2.5 bits of information (per message) in this situation. An increase in the input information cannot be handled by the channel. If the number of dimensions, along which the signal can vary, is increased, the channel capacity increases but at a decreasing rate. Thus the multidimensionality of the stimuli is presumably one of the devices by which human beings can effectively make the multitude of discriminations required in their complex environment.

The "span of immediate memory," which again is in the neighborhood of seven items, seems at first glance to be a rather similar instance of the human observer's limited "channel capacity." However, closer scrutiny of the data shows that in the experiments on absolute judgment, the amount of information is the invariant factor limiting the channel capacity, whereas in the "immediate memory" task the channel capacity seems to be independent of the number of *bits* of information contained in the message, being determined instead by the number of "*chunks*" of information (the number of items constituting the message irrespective of their information content). Subjects can recall about the same number of binary digits, decimal digits, and monosyllabic words, although the amount of information contained in these chunks is vastly different. The limitations imposed on the human observer as a communication channel are quite different in the two situations discussed.

The latter limitation to a fixed number of items in the immediate recall situation can be overcome by grouping or organizing of the many items into new classes which carry more information. As for example in learning Morse code, the input signal containing many chunks of information, with few bits per chunk, is translated into some new code that contains more bits per chunk, but decreases the number of chunks required for the whole message. The input sequence is organized and recoded, usually into a verbal code, and then the verbal code names rather than the original inputs are recalled. After the coding procedure is well learned, the channel capacity (in terms of information transmission) is raised drastically, as is shown in an experiment by S. Smith, in which sequences of binary digits were recoded into octal digits, and the messages recalled were equivalent to about 40 binary digits, as against a span of immediate memory of about 12 binary digits. This recoding process is related to the well-known organized changes in memory, as for example the influence of names on the repro-

duction of memorized visual figures. Information theory has provided some tools for the study of this process which deserves much more attention since as the author points out, "the kind of linguistic recoding that people do seems . . . to be the very life blood of the thought processes." (*F.S.*)

Simon, Herbert A. **Rational choice and the structure of the environment.** *Psychol. Rev.*, 1956, 63, 129–138.

While models of rational behavior especially in economics usually contain complicated choice mechanisms, the adaptive behavior of organisms may not at all require postulation of "maximization" and over-all "utility functions." Simpler "satisficing" mechanisms may be all that is required to meet all the needs of the organism at a specified level. The paper addresses itself to the question: "How simple a set of choice mechanisms can we postulate and still obtain the gross features of observed adaptive choice behavior?"

In dealing with this problem, not only the limited capacities of the organism but also the structural characteristics of the environment must be taken into account. A simple model is developed, in which the environment is considered as a network of paths between points; only at a small randomly distributed fraction of these points can the organism satisfy its needs. The organism itself is assumed to have some storage capacities, so that satisfaction of the need becomes critical only after a specified number of moves. Given an environment, in which purely random movement from point to point would make for a rather small chance of survival, the assumption that the organism can see a few moves ahead would raise its survival chances enormously. This probability of survival depends on four parameters, the proportion of food points, the number of paths at each choice point, the storage capacity of the organism, and the span of its horizon. If the organism has only one need, all that is necessary to assume for rational behavior is that the organism explores at random, looking out for food points, and that it proceeds to eat the food when it sees it.

If the organism has several needs, it requires, besides increased storage capacities, a simple priority mechanism for choosing between two alternative goals. This mechanism might be given by drives proportional to the ratio of the average number of moves necessary to find a goal point to the remaining supply in storage. An even simpler rule would be for the organism to persist in searching for points satisfying the need that first reached the threshold; although this mechanism may not be efficient in respect to the goal-reaching time, even this simplest mechanism may be satisfactory for organisms with sufficient storage capacity. If the environment, in addition to food points, also contains clues indicating the location of such points beyond the horizon, simple conditioning of the priority mechanism to these clues would be all that is required.

The important conclusion drawn from this model is that the allocation of time, the one rare commodity, to the search related to different needs can be accomplished without requiring complicated choice mechanisms or a general utility function. Of course, more efficient mechanisms can be postulated, but for an organism attempting only to "satisfice" but not to "maximize," such postulates may be superfluous and, as consideration of the postulates of economic theories of rational behavior shows, unrealistic. Conflict produced experimentally by forcing the organism to attend simultaneously to alternative goals may only indicate that the structure of the environment does not normally create such situations and the organism therefore is not programmed to deal with such choices. (*F.S.*)

TOWARD A THEORY OF SCHIZOPHRENIA

by Gregory Bateson, Don D. Jackson, Jay Haley, and John Weakland

Veterans Administration Hospital, Palo Alto, California; and Stanford University

Schizophrenia—its nature, etiology, and the kind of therapy to use for it—remains one of the most puzzling of the mental illnesses. The theory of schizophrenia presented here is based on communications analysis, and specifically on the Theory of Logical Types. From this theory and from observations of schizophrenic patients is derived a description, and the necessary conditions for, a situation called the "double bind"—a situation in which no matter what a person does, he "can't win." It is hypothesized that a person caught in the double bind may develop schizophrenic symptoms. How and why the double bind may arise in a family situation is discussed, together with illustrations from clinical and experimental data.

THIS is a report[1] on a research project which has been formulating and testing a broad, systematic view of the nature, etiology, and therapy of schizophrenia. Our research in this field has proceeded by discussion of a varied body of data and ideas, with all of us contributing according to our varied experience in anthropology, communications analysis, psychotherapy, psychiatry, and psychoanalysis. We have now reached common agreement on the broad outlines of a communicational theory of the origin and nature of schizophrenia; this paper is a preliminary report on our continuing research.

[1] This paper derives from hypotheses first developed in a research project financed by the Rockefeller Foundation from 1952–54, administered by the Department of Sociology and Anthropology at Stanford University and directed by Gregory Bateson. Since 1954 the project has continued, financed by the Josiah Macy, Jr. Foundation. To Jay Haley is due credit for recognizing that the symptoms of schizophrenia are suggestive of an inability to discriminate the Logical Types, and this was amplified by Bateson who added the notion that the symptoms and etiology could be formally described in terms of a double bind hypothesis. The hypothesis was communicated to D. D. Jackson and found to fit closely with his ideas of family homeostasis. Since then Dr. Jackson has worked closely with the project. The study of the formal analogies between hypnosis and schizophrenia has been the work of John H. Weakland and Jay Haley.

THE BASE IN COMMUNICATIONS THEORY

Our approach is based on that part of communications theory which Russell has called the Theory of Logical Types (17). The central thesis of this theory is that there is a discontinuity between a class and its members. The class cannot be a member of itself nor can one of the members *be* the class, since the term used for the class is of a *different level of abstraction*—a different Logical Type—from terms used for members. Although in formal logic there is an attempt to maintain this discontinuity between a class and its members, we argue that in the psychology of real communications this discontinuity is continually and inevitably breached (2), and that a priori we must expect a pathology to occur in the human organism when certain formal patterns of the breaching occur in the communication between mother and child. We shall argue that this pathology at its extreme will have symptoms whose formal characteristics would lead the pathology to be classified as a schizophrenia.

Illustrations of how human beings handle

communication involving multiple Logical Types can be derived from the following fields:

1. *The use of various communicational modes in human communication.* Examples are play, non-play, fantasy, sacrament, metaphor, etc. Even among the lower mammals there appears to be an exchange of signals which identify certain meaningful behavior as "play," etc.[2] These signals are evidently of higher Logical Type than the messages they classify. Among human beings this framing and labeling of messages and meaningful actions reaches considerable complexity, with the peculiarity that our vocabulary for such discrimination is still very poorly developed, and we rely preponderantly upon nonverbal media of posture, gesture, facial expression, intonation, and the context for the communication of these highly abstract, but vitally important, labels.

2. *Humor.* This seems to be a method of exploring the implicit themes in thought or in a relationship. The method of exploration involves the use of messages which are characterized by a condensation of Logical Types or communicational modes. A discovery, for example, occurs when it suddenly becomes plain that a message was not only metaphoric but also more literal, or vice versa. That is to say, the explosive moment in humor is the moment when the labeling of the mode undergoes a dissolution and resynthesis. Commonly, the punch line compels a re-evaluation of earlier signals which ascribed to certain messages a particular mode (e.g., literalness or fantasy). This has the peculiar effect of attributing *mode* to those signals which had previously the status of that higher Logical Type which classifies the modes.

3. *The falsification of mode-identifying signals.* Among human beings mode identifiers can be falsified, and we have the artificial laugh, the manipulative simulation of friendliness, the confidence trick, kidding, and the like. Similar falsifications have been recorded among mammals (3, 13). Among human beings we meet with a strange phenomenon—the unconscious falsification of these signals. This may occur within the self—the subject may conceal from himself his own real hostility under the guise of metaphoric play—or it may occur as an unconscious falsification of the subject's understanding of the other person's mode-identifying signals. He may mistake shyness for contempt, etc. Indeed most of the errors of self-reference fall under this head.

4. *Learning.* The simplest level of this phenomenon is exemplified by a situation in which a subject receives a message and acts appropriately on it: "I heard the clock strike and knew it was time for lunch. So I went to the table." In learning experiments the analogue of this sequence of events is observed by the experimenter and commonly treated as a single message of a higher type. When the dog salivates between buzzer and meat powder, this sequence is accepted by the experimenter as a message indicating that "the dog has *learned* that buzzer means meat powder." But this is not the end of the hierarchy of types involved. The experimental subject may become more skilled in learning. He may *learn to learn* (1, 7, 9), and it is not inconceivable that still higher orders of learning may occur in human beings.

5. *Multiple levels of learning and the Logical Typing of signals.* These are two inseparable sets of phenomena—inseparable because the ability to handle the multiple types of signals is itself a *learned* skill and therefore a function of the multiple levels of learning.

According to our hypothesis, the term "ego function" (as this term is used when a schizophrenic is described as having "weak ego function") is precisely *the process of discriminating communicational modes either within the self or between the self and others.* The schizophrenic exhibits weakness in three areas of such function: (*a*) He has difficulty in assigning the correct communicational mode to the messages he receives from other persons. (*b*) He has difficulty in assigning the correct communicational mode to those messages which he himself utters or emits nonverbally. (*c*) He has difficulty in assigning the correct communicational mode

[2] A film prepared by this project, "The Nature of Play; Part I, River Otters," is available.

to his own thoughts, sensations, and percepts.

At this point it is appropriate to compare what was said in the previous paragraph with von Domarus' (16) approach to the systematic description of schizophrenic utterance. He suggests that the messages (and thought) of the schizophrenic are deviant in syllogistic structure. In place of structures which derive from the syllogism, Barbara, the schizophrenic, according to this theory, uses structures which identify predicates. An example of such a distorted syllogism is:

Men die.
Grass dies.
Men are grass.

But as we see it, Von Domarus' formulation is only a more precise—and therefore valuable—way of saying that schizophrenic utterance is rich in metaphor. With that generalization we agree. But metaphor is an indispensable tool of thought and expression—a characteristic of all human communication, even of that of the scientist. The conceptual models of cybernetics and the energy theories of psychoanalysis are, after all, only labeled metaphors. The peculiarity of the schizophrenic is not that he uses metaphors, but that he uses *unlabeled* metaphors. He has special difficulty in handling signals of that class whose members assign Logical Types to other signals.

If our formal summary of the symptomatology is correct and if the schizophrenia of our hypothesis is essentially a result of family interaction, it should be possible to arrive a priori at a formal description of these sequences of experience which would induce such a symptomatology. What is known of learning theory combines with the evident fact that human beings use *context* as a guide for mode discrimination. Therefore, we must look not for some specific traumatic experience in the infantile etiology but rather for characteristic sequential patterns. The specificity for which we search is to be at an abstract or formal level. The sequences must have this characteristic: that from them the patient will acquire the mental habits which are exemplified in schizophrenic communication. That is to say, *he must live in a universe where the sequences of events are such that his unconventional communicational habits will be in some sense appropriate.* The hypothesis which we offer is that sequences of this kind in the external experience of the patient are responsible for the inner conflicts of Logical Typing. For such unresolvable sequences of experiences, we use the term "double bind."

The double bind

The necessary ingredients for a double bind situation, as we see it, are:

1. *Two or more persons.* Of these, we designate one, for purposes of our definition, as the "victim." We do not assume that the double bind is inflicted by the mother alone, but that it may be done either by mother alone or by some combination of mother, father, and/or siblings.

2. *Repeated experience.* We assume that the double bind is a recurrent theme in the experience of the victim. Our hypothesis does not invoke a single traumatic experience, but such repeated experience that the double bind structure comes to be an habitual expectation.

3. *A primary negative injunction.* This may have either of two forms: (*a*) Do not do so and so, or I will punish you," or (*b*) "If you do not do so and so, I will punish you." Here we select a context of learning based on avoidance of punishment rather than a context of reward seeking. There is perhaps no formal reason for this selection. We assume that the punishment may be either the withdrawal of love or the expression of hate or anger—or most devastating—the kind of abandonment that results from the parent's expression of extreme helplessness.[3]

4. A *secondary injunction conflicting with the first at a more abstract level, and like the first enforced by punishments or signals which threaten survival.* This secondary injunction is more difficult to describe than the primary for two reasons. First, the secondary injunction is commonly communicated to the child by nonverbal means.

[3] Our concept of punishment is being refined at present. It appears to us to involve perceptual experience in a way that cannot be encompassed by the notion of "trauma."

Posture, gesture, tone of voice, meaningful action, and the implications concealed in verbal comment may all be used to convey this more abstract message. Second, the secondary injunction may impinge upon any element of the primary prohibition. Verbalization of the secondary injunction may, therefore, include a wide variety of forms; for example, "Do not see this as punishment"; "Do not see me as the punishing agent"; "Do not submit to my prohibitions"; "Do not think of what you must not do"; "Do not question my love of which the primary prohibition is (or is not) an example"; and so on. Other examples become possible when the double bind is inflicted not by one individual but by two. For example, one parent may negate at a more abstract level the injunctions of the other.

5. *A tertiary negative injunction prohibiting the victim from escaping from the field.* In a formal sense it is perhaps unnecessary to list this injunction as a separate item since the reinforcement at the other two levels involves a threat to survival, and if the double binds are imposed during infancy, escape is naturally impossible. However, it seems that in some cases the escape from the field is made impossible by certain devices which are not purely negative, e.g., capricious promises of love, and the like.

6. Finally, the complete set of ingredients is no longer necessary when the victim has learned to perceive his universe in double bind patterns. Almost any part of a double bind sequence may then be sufficient to precipitate panic or rage. The pattern of conflicting injunctions may even be taken over by hallucinatory voices (14).

The effect of the double bind

In the Eastern religion, Zen Buddhism, the goal is to achieve Enlightenment. The Zen Master attempts to bring about enlightenment in his pupil in various ways. One of the things he does is to hold a stick over the pupil's head and say fiercely, "If you say this stick is real, I will strike you with it. If you say this stick is not real, I will strike you with it. If you don't say anything, I will strike you with it." We feel that the schizophrenic finds himself continually in the same situation as the pupil, but he achieves something like disorientation rather than enlightenment. The Zen pupil might reach up and take the stick away from the Master—who might accept this response, but the schizophrenic has no such choice since with him there is no not caring about the relationship, and his mother's aims and awareness are not like the Master's.

We hypothesize that there will be a breakdown in any individual's ability to discriminate between Logical Types whenever a double bind situation occurs. The general characteristics of this situation are the following:

1. When the individual is involved in an intense relationship; that is, a relationship in which he feels it is vitally important that he discriminate accurately what sort of message is being communicated so that he may respond appropriately.

2. And, the individual is caught in a situation in which the other person in the relationship is expressing two orders of message and one of these denies the other.

3. And, the individual is unable to comment on the messages being expressed to correct his discrimination of what order of message to respond to, i.e., he cannot make a metacommunicative statement.

We have suggested that this is the sort of situation which occurs between the preschizophrenic and his mother, but it also occurs in normal relationships. When a person is caught in a double bind situation, he will respond defensively in a manner similar to the schizophrenic. An individual will take a metaphorical statement literally when he is in a situation where he must respond, where he is faced with contradictory messages, and when he is unable to comment on the contradictions. For example, one day an employee went home during office hours. A fellow employee called him at his home, and said lightly, "Well, how did you get *there*?" The employee replied, "By automobile." He responded literally because he was faced with a message which asked him what he was doing at home when he should have been at the office, but which denied that this question was being

asked by the way it was phrased. (Since the speaker felt it wasn't really his business, he spoke metaphorically.) The relationship was intense enough so that the victim was in doubt how the information would be used, and he therefore responded literally. This is characteristic of anyone who feels "on the spot," as demonstrated by the careful literal replies of a witness on the stand in a court trial. The schizophrenic feels so terribly on the spot at all times that he habitually responds with a defensive insistence on the literal level when it is quite inappropriate, e.g., when someone is joking.

Schizophrenics also confuse the literal and metaphoric in their own utterance when they feel themselves caught in a double bind. For example, a patient may wish to criticize his therapist for being late for an appointment, but he may be unsure what sort of a message that act of being late was—particularly if the therapist has anticipated the patient's reaction and apologized for the event. The patient cannot say, "Why were you late? Is it because you don't want to see me today?" This would be an accusation, and so he shifts to a metaphorical statement. He may then say, "I knew a fellow once who missed a boat, his name was Sam and the boat almost sunk, . . . etc.," Thus he develops a metaphorical story and the therapist may or may not discover in it a comment on his being late. The convenient thing about a metaphor is that it leaves it up to the therapist (or mother) to see an accusation in the statement if he chooses, or to ignore it if he chooses. Should the therapist accept the accusation in the metaphor, then the patient can accept the statement he has made about Sam as metaphorical. If the therapist points out that this doesn't sound like a true statement about Sam, as a way of avoiding the accusation in the story, the patient can argue that there really was a man named Sam. As an answer to the double bind situation, a shift to a metaphorical statement brings safety. However, it also prevents the patient from making the accusation he wants to make. But instead of getting over his accusation by indicating that this is a metaphor, the schizophrenic patient seems to try to get over the fact that it is a metaphor by making it more fantastic. If the therapist should ignore the accusation in the story about Sam, the schizophrenic may then tell a story about going to Mars in a rocket ship as a way of putting over his accusation. The indication that it is a metaphorical statement lies in the fantastic aspect of the metaphor, not in the signals which usually accompany metaphors to tell the listener that a metaphor is being used.

It is not only safer for the victim of a double bind to shift to a metaphorical order of message, but in an impossible situation it is better to shift and become somebody else, or shift and insist that he is somewhere else. Then the double bind cannot work on the victim, because it isn't he and besides he is in a different place. In other words, the statements which show that a patient is disoriented can be interpreted as ways of defending himself against the situation he is in. The pathology enters when the victim himself either does not know that his responses are metaphorical or cannot say so. To recognize that he was speaking metaphorically he would need to be aware that he was defending himself and therefore was afraid of the other person. To him such an awareness would be an indictment of the other person and therefore provoke disaster.

If an individual has spent his life in the kind of double bind relationship described here, his way of relating to people after a psychotic break would have a systematic pattern. First, he would not share with normal people those signals which accompany messages to indicate what a person means. His metacommunicative system—the communications about communication—would have broken down, and he would not know what kind of message a message was. If a person said to him, "what would you like to do today?" he would be unable to judge accurately by the context or by the tone of voice or gesture whether he was being condemned for what he did yesterday, or being offered a sexual invitation, or just what was meant. Given this inability to judge accurately what a person really means and an excessive concern with what is really meant, an individual might defend himself by choosing one or more of several alterna-

lives. He might, for example, assume that behind every statement there is a concealed meaning which is detrimental to his welfare. He would then be excessively concerned with hidden meanings and determined to demonstrate that he could not be deceived —as he had been all his life. If he chooses this alternative, he will be continually searching for meanings behind what people say and behind chance occurrences in the environment, and he will be characteristically suspicious and defiant.

He might choose another alternative, and tend to accept literally everything people say to him; when their tone or gesture or context contradicted what they said, he might establish a pattern of laughing off these metacommunicative signals. He would give up trying to discriminate between levels of message and treat all messages as unimportant or to be laughed at.

If he didn't become suspicious of metacommunicative messages or attempt to laugh them off, he might choose to try to ignore them. Then he would find it necessary to see and hear less and less of what went on around him, and do his utmost to avoid provoking a response in his environment. He would try to detach his interest from the external world and concentrate on his own internal processes and, therefore, give the appearance of being a withdrawn, perhaps mute, individual.

This is another way of saying that if an individual doesn't know what sort of message a message is, he may defend himself in ways which have been described as paranoid, hebephrenic, or catatonic. These three alternatives are not the only ones. The point is that he cannot choose the one alternative which would help him to discover what people mean; he cannot, without considerable help, discuss the messages of others. Without being able to do that, the human being is like any self-correcting system which has lost its governor; it spirals into never-ending, but always systematic, distortions.

A DESCRIPTION OF THE FAMILY SITUATION

The theoretical possibility of double bind situations stimulated us to look for such communication sequences in the schizophrenic patient and in his family situation. Toward this end we have studied the written and verbal reports of psychotherapists who have treated such patients intensively; we have studied tape recordings of psychotherapeutic interviews, both of our own patients and others; we have interviewed and taped parents of schizophrenics; we have had two mothers and one father participate in intensive psychotherapy; and we have interviewed and taped parents and patients seen conjointly.

On the basis of these data we have developed a hypothesis about the family situation which ultimately leads to an individual suffering from schizophrenia. This hypothesis has not been statistically tested; it selects and emphasizes a rather simple set of interactional phenomena and does not attempt to describe comprehensively the extraordinary complexity of a family relationship.

We hypothesize that the family situation of the schizophrenic has the following general characteristics:

1. A child whose mother becomes anxious and withdraws if the child responds to her as a loving mother. That is, the child's very existence has a special meaning to the mother which arouses her anxiety and hostility when she is in danger of intimate contact with the child.

2. A mother to whom feelings of anxiety and hostility toward the child are not acceptable, and whose way of denying them is to express overt loving behavior to persuade the child to respond to her as a loving mother and to withdraw from him if he does not. "Loving behavior" does not necessarily imply "affection"; it can, for example, be set in a framework of doing the proper thing, instilling "goodness," and the like.

3. The absence of anyone in the family, such as a strong and insightful father, who can intervene in the relationship between the mother and child and support the child in the face of the contradictions involved.

Since this is a formal description we are not specifically concerned with why the mother feels this way about the child, but we suggest that she could feel this way for various reasons. It may be that merely hav-

ing a child arouses anxiety about herself and her relationships to her own family; or it may be important to her that the child is a boy or a girl, or that the child was born on the anniversary of one of her own siblings (8), or the child may be in the same sibling position in the family that she was, or the child may be special to her for other reasons related to her own emotional problems.

Given a situation with these characteristics, we hypothesize that the mother of a schizophrenic will be simultaneously expressing at least two orders of message. (For simplicity in this presentation we shall confine ourselves to two orders.) These orders of message can be roughly characterized as (*a*) hostile or withdrawing behavior which is aroused whenever the child approaches her, and (*b*) simulated loving or approaching behavior which is aroused when the child responds to her hostile and withdrawing behavior, as a way of denying that she is withdrawing. Her problem is to control her anxiety by controlling the closeness and distance between herself and her child. To put this another way, if the mother begins to feel affectionate and close to her child, she begins to feel endangered and must withdraw from him; but she cannot accept this hostile act and to deny it must simulate affection and closeness with her child. The important point is that her loving behavior is then a comment on (since it is compensatory for) her hostile behavior and consequently it is of a different *order* of message than the hostile behavior—it is a message about a sequence of messages. Yet by its nature it denies the existence of those messages which it is about, i.e., the hostile withdrawal.

The mother uses the child's responses to affirm that her behavior is loving, and since the loving behavior is simulated, the child is placed in a position where he must not accurately interpret her communication if he is to maintain his relationship with her. In other words, he must not discriminate accurately between orders of message, in this case the difference between the expression of simulated feelings (one Logical Type) and real feelings (another Logical Type). As a result the child must systematically distort his perception of metacommunicative signals. For example, if mother begins to feel hostile (or affectionate) toward her child and also feels compelled to withdraw from him, she might say, "Go to bed, you're very tired and I want you to get your sleep." This overtly loving statement is intended to deny a feeling which could be verbalized as "Get out of my sight because I'm sick of you." If the child correctly discriminates her metacommunicative signals, he would have to face the fact that she both doesn't want him and is deceiving him by her loving behavior. He would be "punished" for learning to discriminate orders of messages accurately. He therefore would tend to accept the idea that he is tired rather than recognize his mother's deception. This means that he must deceive himself about his own internal state in order to support mother in her deception. To survive with her he must falsely discriminate his own internal messages as well as falsely discriminate the messages of others.

The problem is compounded for the child because the mother is "benevolently" defining for him how he feels; she is expressing overt maternal concern over the fact that he is tired. To put it another way, the mother is controlling the child's definitions of his own messages, as well as the definition of his responses to her (e.g., by saying, "You don't really mean to say that," if he should criticize her) by insisting that she is not concerned about herself but only about him. Consequently, the easiest path for the child is to accept mother's simulated loving behavior as real, and his desires to interpret what is going on are undermined. Yet the result is that the mother is withdrawing from him and defining this withdrawal as the way a loving relationship should be.

However, accepting mother's simulated loving behavior as real also is no solution for the child. Should he make this false discrimination, he would approach her; this move toward closeness would provoke in her feelings of fear and helplessness, and she would be compelled to withdraw. But if he then withdrew from her, she would take his withdrawal as a statement that she was not a loving mother and would either

punish him for withdrawing or approach him to bring him closer. If he then approached, she would respond by putting him at a distance. *The child is punished for discriminating accurately what she is expressing, and he is punished for discriminating inaccurately—he is caught in a double bind.*

The child might try various means of escaping from this situation. He might, for example, try to lean on his father or some other member of the family. However, from our preliminary observations we think it is likely that the fathers of schizophrenics are not substantial enough to lean on. They are also in the awkward position where if they agreed with the child about the nature of mother's deceptions, they would need to recognize the nature of their own relationships to the mother, which they could not do and remain attached to her in the *modus operandi* they have worked out.

The need of the mother to be wanted and loved also prevents the child from gaining support from some other person in the environment, a teacher, for example. A mother with these characteristics would feel threatened by any other attachment of the child and would break it up and bring the child back closer to her with consequent anxiety when the child became dependent on her.

The only way the child can really escape from the situation is to comment on the contradictory position his mother has put him in. However, if he did so, the mother would take this as an accusation that she is unloving and both punish him and insist that his perception of the situation is distorted. By preventing the child from talking about the situation, the mother forbids him using the metacommunicative level—the level we use to correct our perception of communicative behavior. The ability to communicate about communication, to comment upon the meaningful actions of oneself and others, is essential for successful social intercourse. In any normal relationship there is a constant interchange of metacommunicative messages such as "What do you mean?" or "Why did you do that?" or "Are you kidding me?" and so on. To discriminate accurately what people are really expressing we must be able to comment directly or indirectly on that expression. This metacommunicative level the schizophrenic seems unable to use successfully (2). Given these characteristics of the mother, it is apparent why. If she is denying one order of message, then any statement about her statements endangers her and she must forbid it. Therefore, the child grows up unskilled in his ability to communicate about communication and, as a result, unskilled in determining what people really mean and unskilled in expressing what he really means, which is essential for normal relationships.

In summary, then, we suggest that the double bind nature of the family situation of a schizophrenic results in placing the child in a position where if he responds to his mother's simulated affection her anxiety will be aroused and she will punish him (or insist, to protect herself, that *his* overtures are simulated, thus confusing him about the nature of his own messages) to defend herself from closeness with him. Thus the child is blocked off from intimate and secure associations with his mother. However, if he does not make overtures of affection, she will feel that this means she is not a loving mother and her anxiety will be aroused. Therefore, she will either punish him for withdrawing or make overtures toward the child to insist that he demonstrate that he loves her. If he then responds and shows her affection, she will not only feel endangered again, but she may resent the fact that she had to force him to respond. In either case in a relationship, the most important in his life and the model for all others, he is punished if he indicates love and affection and punished if he does not; and his escape routes from the situation, such as gaining support from others, are cut off. This is the basic nature of the double bind relationship between mother and child. This description has not depicted, of course, the more complicated interlocking gestalt that is the "family" of which the "mother" is one important part (11, 12).

ILLUSTRATIONS FROM CLINICAL DATA

An analysis of an incident occurring between a schizophrenic patient and his mother illustrates the "double bind" situation. A

young man who had fairly well recovered from an acute schizophrenic episode was visited in the hospital by his mother. He was glad to see her and impulsively put his arm around her shoulders, whereupon she stiffened. He withdrew his arm and she asked, "Don't you love me any more?" He then blushed, and she said, "Dear, you must not be so easily embarrassed and afraid of your feelings." The patient was able to stay with her only a few minutes more and following her departure he assaulted an aide and was put in the tubs.

Obviously, this result could have been avoided if the young man had been able to say, "Mother, it is obvious that you become uncomfortable when I put my arm around you, and that you have difficulty accepting a gesture of affection from me." However, the schizophrenic patient doesn't have this possibility open to him. His intense dependency and training prevents him from commenting upon his mother's communicative behavior, though she comments on his and forces him to accept and to attempt to deal with the complicated sequence. The complications for the patient include the following:

1. The mother's reaction of not accepting her son's affectionate gesture is masterfully covered up by her condemnation of him for withdrawing, and the patient denies his perception of the situation by accepting her condemnation.

2. The statement "don't you love me any more" in this context seems to imply:

(*a*) "I am lovable."

(*b*) "You should love me and if you don't you are bad or at fault."

(*c*) "Whereas you did love me previously you don't any longer," and thus focus is shifted from his expressing affection to his inability to be affectionate. Since the patient has also hated her, she is on good ground here, and he responds appropriately with guilt, which she then attacks.

(*d*) "What you just expressed *was not* affection," and in order to accept this statement the patient must deny what she and the culture have taught him about how one expresses affection. He must also question the times with her, and with others, when he thought he was experiencing affection and when they *seemed* to treat the situation as if he had. He experiences here loss-of-support phenomena and is put in doubt about the reliability of past experience.

3. The statement, "You must not be so easily embarrassed and afraid of your feelings," seems to imply;

(*a*) "You are not like me and are different from other nice or normal people because we express our feelings."

(*b*) "The feelings you express are all right, it's only that *you* can't accept them." However, if the stiffening on her part had indicated "these are unacceptable feelings," then the boy is told that he should not be embarrassed by unacceptable feelings. Since he has had a long training in what is and is not acceptable to both her and society, he again comes into conflict with the past. If he is unafraid of his own feelings (which mother implies is good), he should be unafraid of his affection and would then notice it was she who was afraid, but he must not notice that because her whole approach is aimed at covering up this shortcoming in herself.

The impossible dilemma thus becomes: "If I am to keep my tie to mother I must not show her that I love her, but if I do not show her that I love her, then I will lose her."

The importance to the mother of her special method of control is strikingly illustrated by the interfamily situation of a young woman schizophrenic who greeted the therapist on their first meeting with the remark, "Mother had to get married and now I'm here." This statement meant to the therapist that:

1. The patient was the result of an illegitimate pregnancy.

2. This fact was related to her present psychosis (in her opinion).

3. "Here" referred to the psychiatrist's office and to the patient's presence on earth for which she had to be eternally indebted to her mother, especially since her mother had sinned and suffered in order to bring her into the world.

4. "Had to get married" referred to the shot-gun nature of mother's wedding and to

the mother's response to pressure that she must marry, and the reciprocal, that she resented the forced nature of the situation and blamed the patient for it.

Actually, all these suppositions subsequently proved to be factually correct and were corroborated by the mother during an abortive attempt at psychotherapy. The flavor of the mother's communications to the patient seemed essentially this: "I am lovable, loving, and satisfied with myself. You are lovable when you are like me and when you do what I say." At the same time the mother indicated to the daughter both by words and behavior: "You are physically delicate, unintelligent, and different from me ('not normal'). You need me and me alone because of these handicaps, and I will take care of you and love you." Thus the patient's life was a series of beginnings, of attempts at experience, which would result in failure and withdrawal back to the maternal hearth and bosom because of the collusion between her and her mother.

It was noted in collaborative therapy that certain areas important to the mother's self-esteem were especially conflictual situations for the patient. For example, the mother needed the fiction that she was close to her family and that a deep love existed between her and her own mother. By analogy the relationship to the grandmother served as the prototype for the mother's relationship to her own daughter. On one occasion when the daughter was seven or eight years old the grandmother in a rage threw a knife which barely missed the little girl. The mother said nothing to the grandmother but hurried the little girl from the room with the words, "Grandmommy really loves you." It is significant that the grandmother took the attitude toward the patient that she was not well enough controlled, and she used to chide her daughter for being too easy on the child. The grandmother was living in the house during one of the patient's psychotic episodes, and the girl took great delight in throwing various objects at the mother and grandmother while they cowered in fear.

Mother felt herself very attractive as a girl, and she felt that her daughter resembled her rather closely, although by damning with faint praise it was obvious that she felt the daughter definitely ran second. One of the daughter's first acts during a psychotic period was to announce to her mother that she was going to cut off all her hair. She proceeded to do this while the mother pleaded with her to stop. Subsequently the mother would show a picture of *herself* as a girl and explain to people how the patient would look if she only had her beautiful hair.

The mother, apparently without awareness of the significance of what she was doing, would equate the daughter's illness with not being very bright and with some sort of organic brain difficulty. She would invariably contrast this with her own intelligence as demonstrated by her *own* scholastic record. She treated her daughter with a completely patronizing and placating manner which was insincere. For example, in the psychiatrist's presence she promised her daughter that she would not allow her to have further shock treatments, and as soon as the girl was out of the room she asked the doctor if he didn't feel she should be hospitalized and given electric shock treatments. One clue to this deceptive behavior arose during the mother's therapy. Although the daughter had had three previous hospitalizations the mother had never mentioned to the doctors that she herself had had a psychotic episode when she discovered that she was pregnant. The family whisked her away to a small sanitarium in a nearby town, and she was, according to her own statement, strapped to a bed for six weeks. Her family did not visit her during this time, and no one except her parents and her sister knew that she was hospitalized.

There were two times during therapy when the mother showed intense emotion. One was in relating her own psychotic experience; the other was on the occasion of her last visit when she accused the therapist of trying to drive her crazy by forcing her to choose between her daughter and her husband. Against medical advice, she took her daughter out of therapy.

The father was as involved in the homeostatic aspects of the intrafamily situation as

the mother. For example, he stated that he had to quit his position as an important attorney in order to bring his daughter to an area where competent psychiatric help was available. Subsequently, acting on cues from the patient (e.g., she frequently referred to a character named "Nervous Ned") the therapist was able to elicit from him that he had hated his job and for years had been trying to "get out from under." However, the daughter was made to feel that the move was initiated for her.

On the basis of our examination of the clinical data, we have been impressed by a number of observations including:

1. The helplessness, fear, exasperation, and rage which a double bind situation provokes in the patient, but which the mother may serenely and un-understandingly pass over. We have noted reactions in the father that both create double bind situations, or extend and amplify those created by the mother, and we have seen the father passive and outraged, but helpless, become ensnared in a similar manner to the patient.

2. The psychosis seems, in part, a way of dealing with double bind situations to overcome their inhibiting and controlling effect. The psychotic patient may make astute, pithy, often metaphorical remarks that reveal an insight into the forces binding him. Contrariwise, he may become rather expert in setting double bind situations himself.

3. According to our theory, the communication situation described is essential to the mother's security, and by inference to the family homeostasis. If this be so, then when psychotherapy of the patient helps him become less vulnerable to mother's attempts at control, anxiety will be produced in the mother. Similarly, if the therapist interprets to the mother the dynamics of the situation she is setting up with the patient, this should produce an anxiety response in her. Our impression is that when there is a perduring contact between patient and family (especially when the patient lives at home during psychotherapy), this leads to a disturbance (often severe) in the mother and sometimes in both mother and father and other siblings (10, 11).

CURRENT POSITION AND FUTURE PROSPECTS

Many writers have treated schizophrenia in terms of the most extreme contrast with any other form of human thinking and behavior. While it is an isolable phenomenon, so much emphasis on the differences from the normal—rather like the fearful physical segregation of psychotics—does not help in understanding the problems. In our approach we assume that schizophrenia involves general principles which are important in all communication and therefore many informative similarities can be found in "normal" communication situations.

We have been particularly interested in various sorts of communication which involve both emotional significance and the necessity of discriminating between orders of message. Such situations include play, humor, ritual, poetry, and fiction. Play, especially among animals, we have studied at some length (3). It is a situation which strikingly illustrates the occurrence of metamessages whose correct discrimination is vital to the cooperation of the individuals involved; for example, false discrimination could easily lead to combat. Rather closely related to play is humor, a continuing subject of our research. It involves sudden shifts in Logical Types as well as discrimination of those shifts. Ritual is a field in which unusually real or literal ascriptions of Logical Type are made and defended as vigorously as the schizophrenic defends the "reality" of his delusions. Poetry exemplifies the communicative power of metaphor—even very unusual metaphor—when labeled as such by various signs, as contrasted to the obscurity of unlabeled schizophrenic metaphor. The entire field of fictional communication, defined as the narration or depiction of a series of events with more or less of a label of actuality, is most relevant to the investigation of schizophrenia. We are not so much concerned with the content interpretation of fiction—although analysis of oral and destructive themes is illuminating to the student of schizophrenia—as with the formal problems involved in simultaneous existence of multiple levels of message in

the fictional presentation of "reality." The drama is especially interesting in this respect, with both performers and spectators responding to messages about both the actual and the theatrical reality.

We are giving extensive attention to hypnosis. A great array of phenomena that occur as schizophrenic symptoms—hallucinations, delusions, alterations of personality, amnesias, and so on—can be produced temporarily in normal subjects with hypnosis. These need not be directly suggested as specific phenomena, but can be the "spontaneous" result of an arranged communication sequence. For example, Erickson (4) will produce a hallucination by first inducing catalepsy in a subject's hand and then saying, "There is no conceivable way in which your hand can move, yet when I give the signal, it must move." That is, he tells the subject his hand will remain in place, yet it will move, and in no way the subject can consciously conceive. When Erickson gives the signal, the subject hallucinates the hand moved, or hallucinates himself in a different place and therefore the hand was moved. This use of hallucination to resolve a problem posed by contradictory commands which cannot be discussed seems to us to illustrate the solution of a double bind situation via a shift in Logical Types. Hypnotic responses to direct suggestions or statements also commonly involve shifts in type, as in accepting the words "Here's a glass of water" or "You feel tired" as external or internal reality, or in literal response to metaphorical statements, much like schizophrenics. We hope that further study of hypnotic induction, phenomena, and waking will, in this controllable situation, help sharpen our view of the essential communicational sequences which produce phenomena like those of schizophrenia.

Another Erickson experiment (12) seems to isolate a double bind communicational sequence without the specific use of hypnosis. Erickson arranged a seminar so as to have a young chain smoker sit next to him and to be without cigarettes; other participants were briefed on what to do. All was ordered so that Erickson repeatedly turned to offer the young man a cigarette, but was always interrupted by a question from someone so that he turned away, "inadvertently" withdrawing the cigarettes from the young man's reach. Later another participant asked this young man if he had received the cigarette from Dr. Erickson. He replied, "What cigarette?", showed clearly that he had forgotten the whole sequence, and even refused a cigarette offered by another member, saying that he was too interested in the seminar discussion to smoke. This young man seems to us to be in an experimental situation paralleling the schizophrenic's double bind situation with mother: An important relationship, contradictory messages (here of giving and taking away), and comment blocked—because there was a seminar going on, and anyway it was all "inadvertent." And note the similar outcome: Amnesia for the double bind sequence and reversal from "He doesn't give" to "I don't want."

Although we have been led into these collateral areas, our main field of observation has been schizophrenia itself. All of us have worked directly with schizophrenic patients and much of this case material has been recorded on tape for detailed study. In addition, we are recording interviews held jointly with patients and their families, and we are taking sound motion pictures of mothers and disturbed, presumably preschizophrenic, children. Our hope is that these operations will provide a clearly evident record of the continuing, repetitive double binding which we hypothesize goes on steadily from infantile beginnings in the family situation of individuals who become schizophrenic. This basic family situation, and the overtly communicational characteristics of schizophrenia, have been the major focus of this paper. However, we expect our concepts and some of these data will also be useful in future work on other problems of schizophrenia, such as the variety of other symptoms, the character of the "adjusted state" before schizophrenia becomes manifest, and the nature and circumstances of the psychotic break.

THERAPEUTIC IMPLICATIONS OF THIS HYPOTHESIS

Psychotherapy itself is a context of multilevel communication, with exploration of the ambiguous lines between the literal and metaphoric, or reality and fantasy, and indeed, various forms of play, drama, and hypnosis have been used extensively in therapy. We have been interested in therapy, and in addition to our own data we have been collecting and examining recordings, verbatim transcripts, and personal accounts of therapy from other therapists. In this we prefer exact records since we believe that how a schizophrenic talks depends greatly, though often subtly, on how another person talks to him; it is most difficult to estimate what was really occurring in a therapeutic interview if one has only a description of it, especially if the description is already in theoretical terms.

Except for a few general remarks and some speculation, however, we are not yet prepared to comment on the relation of the double bind to psychotherapy. At present we can only note:

1. Double bind situations are created by and within the psychotherapeutic setting and the hospital milieu. From the point of view of this hypothesis we wonder about the effect of medical "benevolence" on the schizophrenic patient. Since hospitals exist for the benefit of personnel as well as—as much as—more than—for the patient's benefit, there will be contradictions at times in sequences where actions are taken "benevolently" for the patient when actually they are intended to keep the staff more comfortable. We would assume that whenever the system is organized for hospital purposes and it is announced to the patient that the actions are for *his* benefit, then the schizophrenogenic situation is being perpetuated. This kind of deception will provoke the patient to respond to it as a double bind situation, and his response will be "schizophrenic" in the sense that it will be indirect and the patient will be unable to comment on the fact that he feels that he is being deceived. One vignette, fortunately amusing, illustrates such a response. On a ward with a dedicated and "benevolent" physician in charge there was a sign on the physician's door which said "Doctor's Office. Please Knock." The doctor was driven to distraction and finally capitulation by the obedient patient who carefully knocked every time he passed the door.

2. The understanding of the double bind and its communicative aspects may lead to innovations in therapeutic technique. Just what these innovations may be is difficult to say, but on the basis of our investigation we are assuming that double bind situations occur consistently in psychotherapy. At times these are inadvertent in the sense that the therapist is imposing a double bind situation similar to that in the patient's history, or the patient is imposing a double bind situation on the therapist. At other times therapists seem to impose double binds, either deliberately or intuitively, which force the patient to respond differently than he has in the past.

An incident from the experience of a gifted psychotherapist illustrates the intuitive understanding of a double bind communicational sequence. Dr. Frieda Fromm-Reichmann (5) was treating a young woman who from the age of seven had built a highly complex religion of her own replete with powerful Gods. She was very schizophrenic and quite hesitant about entering into a therapeutic situation. At the beginning of the treatment she said, "God R says I shouldn't talk with you." Dr. Fromm-Reichmann replied, "Look, let's get something into the record. To me God R doesn't exist, and that whole world of yours doesn't exist. To you it does, and far be it from me to think that I can take that away from you, I have no idea what it means. So I'm willing to talk with you in terms of that world, if only you know I do it so that we have an understanding that it doesn't exist for me. Now go to God R and tell him that we have to talk and he should give you permission. Also you must tell him that I am a doctor and that you have lived with him in his kingdom now from seven to sixteen—that's nine years—and he hasn't helped you. So now he must permit me to try and see whether you and I

can do that job. Tell him that I am a doctor and this is what I want to try."

The therapist has her patient in a "therapeutic double bind." If the patient is rendered doubtful about her belief in her god then she is agreeing with Dr. Fromm-Reichmann, and is admitting her attachment to therapy. If she insists that God R is real, then she must tell him that Dr. Fromm-Reichmann is "more powerful" than he—again admitting her involvement with the therapist.

The difference between the therapeutic bind and the original double bind situation is in part the fact that the therapist is not involved in a life and death struggle himself. He can therefore set up relatively benevolent binds and gradually aid the patient in his emancipation from them. Many of the uniquely appropriate therapeutic gambits arranged by therapists seem to be intuitive. We share the goal of most psychotherapists who strive toward the day when such strokes of genius will be well enough understood to be systematic and commonplace.

REFERENCES

1. Bateson, G. Social planning and the concept of "deutero-learning". *Conference on Science, Philosophy, and Religion, Second Symposium.* New York: Harper, 1942.
2. Bateson, G. A theory of play and fantasy. *Psychiatric Research Reports*, 1955, 2, 39–51.
3. Carpenter, C. R. A field study of the behavior and social relations of howling monkeys. *Comp. Psychol. Monogr.*, 1934, 10, 1–168.
4. Erickson, M. H. Personal communication, 1955.
5. Fromm-Reichmann, F. Personal communication, 1956.
6. Haley, J. Paradoxes in play, fantasy, and psychotherapy. *Psychiatric Research Reports*, 1955, 2, 52–58.
7. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, 56, 51–65.
8. Hilgard, J. R. Anniversary reactions in parents precipitated by children. *Psychiatry*, 1953, 16, 73–80.
9. Hull, C. L., *et al. Mathematico-deductive theory of rote learning.* New Haven: Yale Univ. Press, 1940.
10. Jackson, D. D. An episode of sleepwalking. *J. Amer. Psychoanal. Assn.*, 1954, 2, 503–508.
11. Jackson, D. D. Some factors influencing the Oedipus complex. *Psychoanal. Quart.*, 1954, 23, 566–581.
12. Jackson, D. D. The question of family homeostasis. Presented at the Amer. Psychiatric Assn. Meeting, St. Louis, May 7, 1954.
13. Lorenz, K. Z. *King Solomon's ring.* New York: Crowell, 1952.
14. Perceval, J. A narrative of the treatment experienced by a gentleman during a state of mental derangement, designed to explain the causes and nature of insanity, etc. London: Effingham Wilson, 1836 and 1840.
15. Ruesch, J., & Bateson, G. *Communication: the social matrix of psychiatry.* New York: Norton, 1951.
16. von Domarus, E. The specific laws of logic in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univ. of California Press, 1944.
17. Whitehead, A. N., & Russell, B. *Principia mathematica.* Cambridge: Cambridge Univ. Press, 1910.

(Manuscript received June 7, 1956)

Those who have read of everything are thought to understand everything too; but it is not always so—reading furnishes the mind only with materials of knowledge; it is thinking that makes what is read ours. We are of the ruminating kind, and it is not enough to cram ourselves with a great load of collections; unless we chew them over again, they will not give us strength and nourishment.

—John Locke

THE LOGICAL NATURE OF AN ACTION SCHEME

by Manfred Kochen and Marion J. Levy, Jr.*

Department of Economics and Sociology, Princeton University

A chronic concern of the behavioral scientist is how to build a system of definitions for the terms he uses so that when two or more behavioral scientists talk, they can be sure they understand each other. Consequently, much of theoretical social science has been little more than an effort to come to terms, to agree how the concepts of social science are to be abstracted and to what observables they are to refer.

"Action" has been one of the central concepts whose sphere of definition various social scientists have tried to circumscribe. The authors examine some of the attempts and propose their own, which, they hope, will result in the emergence of concepts sufficiently precise to allow the development of behavioral science along the lines of inductive-deductive method which has been so successful in natural science.

I. A RESIDUAL DEFINITION OF ACTION

IN THE development of theories in social science, the concept of social action has occupied a dominant role. In distinction to the empirical study of action, which began for present purposes essentially at the start of the twentieth century with such groups as the Würzburg school (1, 2, 3, 13) the notion of an action-language, as a tool for organizing concepts and assumptions, emerged from the German idealistic tradition (21, 29). In recent years, the action-schematization has been refined by Parsons (20). To permit the eventual evaluation of the fruitfulness of Parsons' revival of the German action-language, it is desirable to analyze the basic concepts in a more logically precise manner than has hitherto appeared in the literature. It is the primary objective of this article to clarify the logical nature of an action scheme, and demonstrate its relationship to mathematical models. Thus, the present state of theoretical knowledge concerning action theory is translated into formal language, and its limitations, scope, and structure are exhibited. The relation to other conceptual approaches to the definition of action is studied.

* This author's work was supported by a Fellowship from the Ford Foundation.

The logical nature of an action scheme

It is desired to dissect and study the purely logical aspects of the notion of *Handlung* as a frame of reference within which available knowledge and prevailing procedures can be located. For instance, such a conceptual structure should specify those characteristics of a crime which are sufficient to permit a judicial-legal decision (21) (punishment for doing something, failing to do something, or intending to do something). It is convenient first to introduce the most recent definition of "action" which has been proposed (12, p. 8) and to discuss its relation to the definitions of Parsons and Marschak. In Levy's definition, the meaning of an "operation" is assumed to be understood. Then, the definition of action proceeds by stages:

1. An operation associated with an individual of a given species is called an *action* if there exists an experiment which can determine whether the operation has or has not taken place. For example, the operation of "flying" as performed by a bird can be demonstrated to have taken place by visual observation, hence it is an action. The operation of "communicating with spirits" as "performed" by a mystic cannot be demonstrated to have occurred

by existing methods, hence is not an action. Note that the operation of "believing to have communicated with spirits" is an action, since it is possible to set up means of determining whether such a belief is or is not exhibited.

2. An action is called a *social action* if it cannot be demonstrated to depend (deterministically or probabilistically) on the factors of the heredity and/or the nonspecies environment of the given species by logical and/or empirical means alone, on a specified level of generalization. For example, the heliotropic action of *Porthesia crysorrhoea* (24), is not a social action, because it can be explained on the level of generalization of the interest in the problem in terms of its heredity and nonspecies environment alone.

3. A social action is called a *human social action*, if the species under consideration is homo sapiens.

According to T. Parsons (20) a sample of behavior (undefined) is called an action[1] if:

1. it is associated with an actor who expends energy;

2. it is "oriented" toward the attainment of an "anticipated state of affairs" (i.e., the "end");

3. it is initiated in a "situation," of which the trends of development differ in one or more important respects from the state of affairs to which the action is oriented; the situation consists of:

a) those elements which the actor cannot alter or prevent from being altered, in conformity with his end (i.e. the "conditions");

b) those elements over which he has control (i.e. the "means");

4. it is normatively regulated. To illustrate with an example taken from Parsons, a "man driving his automobile to a lake to go fishing" is an action. The man is the actor associated with the behavior, and driving involves an expenditure of energy. The anticipated state of affairs to which the man's behavior is oriented is "to go fishing." The behavior takes place in a situation, namely the road (conditions) and the car (means), which is different from the "situation" in which the man will be fishing. This behavior is normatively regulated, because he drives on the right side of the road, stops at red lights, etc.

A third definition of action proposed by Jacob Marschak (14) regards an action as a variable which is controlled by the actor, and which, together with the "state of the environment," determines the actor's satisfaction. An action is based upon an observation made by the actor on the external world and a set of messages (which represent the observations of others) so as to maximize his satisfaction.

The above three definitions can serve as a basis for the following discussion of the logical aspects of an action scheme. First, it should be observed that an action in the sense of Marschak seems to correspond to the means of the situation in Parsons' definition; the conditions of the situation for Parsons correspond to the "state of the environment" in Marschak's model. The actor's tendency to maximize his satisfaction in Marschak's model is a special case of the orientation toward the attainment of an anticipated state of affairs of Parsons' actor. The concepts of normative regulation and expenditure of energy do not appear in Marschak's formulation, because Marschak is concerned with finding the most efficient rules of action and communication for group members, not in the description of existing modes of behavior in human groups. Marschak regards the introduction of the economic, normative approach as a second approximation. Hence, with regard to assumptions 1–3 of Parsons, Marschak's action is a special case of Parsons' scheme.

Secondly, with regard to Parsons' action scheme, it does not seem necessary to introduce the assumption that behavior must be accompanied by the expenditure of energy to be an action. It seems reasonable to regard the expenditure of energy as an aspect which is implicit in the notion of behavior.

To determine whether the "trends of development" of a situation do or do not differ from the anticipated state of affairs, it is necessary that the former be comparable with the latter. Otherwise, the question of

[1] For Parsons "action" is the equivalent of "social action" for Levy.

their sameness is meaningless. However, to make them comparable admits the possibility of regarding the "anticipated state of affairs" as those "trends of development of a situation" which maximize the satisfaction to the actor. This is the situation of game and decision theory. If, however, the "anticipated state of affairs" is regarded to include aspects which are not in the description of the "trends of development of a situation" (e.g., ideas about nonempirical states such as "getting to heaven") the two differ in at least those aspects, and are therefore incomparable. Thus, the assumption that the "anticipated state of affairs" differs from the "trends of development of the situation" is redundant. Hence, the basic elements in Parsons' action scheme are the associated actor, the end, the means, the conditions, and the norms.

The consideration of ends as a basic characteristic in the description of an action is equivalent to the consideration of "subjectively" oriented actions, and is associated with the admission of a "voluntaristic" element. In the analysis of a specific action, the detection of this "voluntaristic" element or of a "deliberate goal" requires the elimination of the possibility that the entire action is not simply a biological response. It is the normative element which makes action *social* for Parsons. In order to determine whether or not a given sample of behavior fits a specified set of norms involves the determination of whether this behavior is a consequence of only environmental and hereditary factors or not. In effect, then, to require that action be normatively regulated behavior is to require, in part, that it cannot be adequately analyzed or explained in terms of hereditary and environmental factors alone—making a behavior sample which is an action in Parsons' sense, a human social action in the sense of Levy.

Levy's conception of an action scheme applies to a wider range of behaviors and permits a somewhat more precise determination than Parsons' of whether or not a given behavior sample is action in Parsons' sense. The "actors," "means," and "conditions" associated with behavior are observed just as in Parsons' scheme. However, it is possible to attribute an "end" and "norms" to any behavior that has not been demonstrated to result from environmental-hereditary forces, although there are types of behavior for which the "ends" and "norms" are not evident, which still might qualify as "actions." In such cases as these latter "actions" early classification of them as "actions" might be a function of ignorance about the setting factors (i.e. species heredity and the nonspecies environment). Increase of knowledge in this sphere might reduce these actions either to the category of "behavior" or definitely establish them as examples of "action" in Parsons' sense; e.g., an increase in knowledge of the physiology of "intelligence" might put certain developments of "intelligence" in the "behavior" category and discriminate others as clearly in the "action" category with a clarity and precision rather beyond our reach at present. A classic example of this sort would be the discrimination between the biological and nonbiological factors in the development of musical genius such as that exhibited by Mozart. Furthermore, the present state of knowledge regarding environmental-hereditary factors is possibly more advanced than the basis for substantive analysis in social science. In conclusion, any action in Marschak's sense of the term can be regarded as an action in Parsons' usage (though not conversely), and the latter can be regarded as an action in Levy's concept (though not conversely).

Both Parsons' and Levy's definitions are understood to be relative to the state of knowledge in the social or natural sciences, which varies with time and with the analyst. A purely logical definition, which is to serve the purposes of logical deduction, should be devoid of any such dependences and empirical connotations in its *formal structure*. Nevertheless, if the definition of action is to be, in fact, residual, it is better to make it explicitly as such.

It may prove fruitful to attempt to describe behavior from a very general viewpoint which has been employed by some scientists (10). It is supposed that an organism is in a certain state, the outside world acts as a stimulus, and the organism

responds in a manner which depends on its state and the stimulus. The universe of discourse for the description of a particular action scheme is assumed to consist of several interconnected organisms of this type and the relevant aspects of their environment. The unit organisms are interpreted as the actors. Transitions in a given unit from one state to another, or what may be interpreted as responses of this unit, are accompanied by events, which may produce state transitions in another unit or in the environment. These, in turn, serve as "experiences" for the first actor-unit, and partially determine the change in state. Time sequences of responses and state changes of an actor-unit are identified with behavior samples or "acts" under special conditions.

Relation to mathematical models

There are mathematical models which, upon suitable interpretations of the basic concepts, can be applied to the general action scheme just proposed (7, 9, 11). One such model, or, more accurately, a framework for a mathematical model, will be briefly described, and it will be shown how Marschak's and Parsons' action schemes can be regarded as particular interpretations. It will be possible to state with some precision what is and what is not assumed, defined, or proved, and to what range of empirical situations the applications extend. Nevertheless, the foundations for a deductive scheme have yet to be developed.

We are presumably concerned with actors, their actions, and the determinants of those actions. The determinants may be thought of as being of two kinds, the internal states of the actors and the state of the world. Therefore in a system involving N actors there will be N variables describing the actions and N the internal states of the actors and one variable describing the state of the world. Needless to say, none of these variables need be nor indeed are likely to be simple scalars, or even any kind of quantities with numerical properties. These basic concepts can be formalized as follows:

1. The universe of discourse is, at any of a discrete number of instants of time, $h = 0, 1, \cdots$, assumed to be completely described by a finite number of variables, $x_1^h, x_2^h \cdots x_N^h, y^h, s_1^h, s_2^h \cdots s_N^h$ which may vary with h, the time. Such a description will be called an "analytical" or "functional" description.

2. The universe of discourse is also described "concretely" or "structurally" by specifying the units of which it is composed, and how they are interconnected. If the universe of discourse is denoted by U, the constituents shall be denoted by $P_1, \cdots, P_N$ and E, and they shall be specified as follows: P_1 is described by the two variables x_1^h and s_1^h, which are defined for any h, and which depend functionally or probabilistically (see below) on $x_1^{h-1}, x_2^{h-1}, \cdots, x_N^{h-1}, y^{h-1}, s_1^{h-1}$. Similarly, P_2 is described by the variables x_2^h and s_2^h which depend on $x_1^{h-1}, \cdots, x_N^{h-1}, y^{h-1}, s_2^{h-1}$, etc. until P_N. E is described by y^h which depends on y^{h-1} and $x_1^h, \cdots, x_N^h$.

As already pointed out, the variables mentioned above need not be quantities with numerical properties, although those are not excluded. It is, however, assumed that each of the variables takes on values which are the elements of sets (collection of all possible values), $X_1, \cdots, X_N, Y, S_1, \cdots, S_n$, which are specified a priori. The set X_1 may be interpreted as all the possible "outputs" which could be associated with P_1, the first of the actor-units in the system (universe). For example, X_1 might be the set of 13 cards, one of which can be played at each turn by the first player in a game of bridge. Clearly, the specification of the sets depends upon the particular problems being formally described. It may be remarked that an element of X_1 can also equivalently be viewed as a transition from one "effective state" to another, so that the set of all possible "effective states" X_1' may be specified in place of X_1. For instance, X_1' may be the set of all combinations of cards and voids adding to 13 which may be held by the first bridge player at any time. X_1 would then be 13 of the large number of possible transitions of X_1' into itself.

As another example, if X_1' consists of 3

elements, a, b, c, then X_1 could consist of as many as 9 elements,

$$a \to a, \quad a \to b, \quad a \to c$$
$$b \to a, \quad b \to b, \quad b \to c$$
$$c \to a, \quad c \to b, \quad c \to c.$$

The set S_1 may be interpreted as the set of all possible "internal states" in which P_1 could be at any time.[2] For example, S_1 might be the set consisting of all considered goals (e.g., to acquire just enough points to win a rubber, to prevent the opponent from winning, to convey information to his partner, to save face, etc.), all considered states of proficiency and information, all considered norms (e.g., not to cheat, to respond quickly, etc.) which are relevant to the first bridge player. If all these possibilities are itemized, each element in S_1, the totality of such possibilities, corresponds to an integer; the values of the variable s_1^h are these integers, but no numerical properties are attributed to s_1^h (e.g., $1 + 2$ is not defined). The set Y is the set of all possible "states of nature," or the environment external to $P_1, \cdots, P_N$. For instance, Y may be the set of all possible sounds, odors, air pressures, temperatures, etc. in the room where the bridge game is being played.

The total state of P_1 at the h^{th} instant of time (since the beginning of observations) is described by x_1^h and s_1^h, elements of X_1 and S_1 or by $x_1'^h$ and s_1^h. An "act" by P_1 at time h is either a transition $x_1'^h \to x_1'^{h+1}$ or x_1^h, depending upon whether $X_1'^h$ or X_1^h is used to describe the "output" of P_1. The transitions $x_1'^h \to x_1'^{h+1}$ may be regarded as involving events which may partially determine events associated with the other variables. It is assumed that *the transitions* $s_1^h \to s_1^{h+1}$ have no appreciable effects on the other variables. *It is, however, assumed that there exist some conceptual experiments by means of which any value of* s_1^h *could be observed.* Note that the set X_1 is the collection of some of the mappings of the set X_1' into itself, and suppose, for concreteness, that the sets $X_1, \cdots, X_N$ rather than $X_1', \cdots, X_N'$ are specified.

The components $P_1, \cdots, P_N, E$ of U may be regarded as deterministic or as probabilistic elements. In the first case, it is assumed that for every P, say for P_1, two functions are given:[3]

$$\begin{aligned} x_1^{h+1} &= f_1(x_1^h, \cdots, x_N^h, s_1^h, y^h) \\ s_1^{h+1} &= g_1(x_1^h, \cdots, x_N^h, s_1^h, y^h); \end{aligned} \qquad [1]$$

and for E, the function

$$y^{h+1} = F(x_1^{h+1}, \cdots, x_N^{h+1}, y^h)$$

is given. In the second case, the joint conditional probabilities

$$\begin{aligned} &Pr\{x_1^{h+1}, s_1^{h+1} \mid x_1^h, \cdots, x_N^h, s_1^h, y^h\}; \\ &\qquad Pr\{y^{h+1} \mid x_1^{h+1}, \cdots, x_N^{h+1}, y^h\} \end{aligned} \qquad [2]$$

are assumed to be given. The functions for $P_1, \cdots, P_N$ are called the behavior functions of the actor-units. Observe that the internal state variables are singled out by appearing in the argument of the behavior function for only one unit. In certain applications, it may be more appropriate to regard time as continuous, so that the behavior functions are written as differential rather than as difference equations. This can be done if X, Y, S are sets of real numbers. Finally, it may be necessary or convenient to regard the internal state of P_1 as described by more than one, say by m_1 variables, $s_{1,1}^h, \cdots, s_{1,m_1}^h$. The sets of which their values are elements, $S_{1,1}, \cdots, S_{1,m_1}$, will be called the factors of the internal state of P_1. Thus, $S_{1,1}$ may be the set of all goals or anticipated states of affairs for P_1, $S_{1,2}$ the set of all norms to which the actions of P_1 might be subject, $S_{1,3}$ the set of all scores P_1 might make on an IQ test, $S_{1,4}$ those he might make on a personality test, $S_{1,5}$ might be the scores P_1 could obtain on an achievement test, etc. Similarly, more than one type of "output"

[2] In another paper by the authors, a distinction is made between the factor S_1 which is interpreted as above, and a corresponding factor Σ_1, which is interpreted as item of information about S_1 which is stored by P_1. It represents P_1's recollection of the experience or perception associated with S_1.

[3] A function is an assignment of a value of a specified set (the range) to every possible argument.

might be assigned to P_1. Thus, instead of one set of all possible acts which P_1 could make in a time interval, M_1 such sets could be considered; i.e., $X_{1,1}$ = set of all sounds made by P_1 in, say, 1 second; $X'_{1,2}$ = set of all positions of the right hand (say at tip of thumb) of P_1 at an instant, so that $X_{1,2}$ = set of all right-hand motions P_1 could have made in one second = mapping of $X'_{1,2}$ into itself, etc. In the case of E, it will almost surely be useful to consider more than the single set Y, say $Y_1, \cdots, Y_n$. Thus Y_1 might be the set of all pressures, Y_2 = temperature range, Y_3 = humidity range, Y_4 = brightness, etc. These sets, Y_1, $Y_2, \cdots$ will be called the factors of the natural environment of any P_1 while all other factors except $S_{1,1} \cdots S_{1,m_1}$, $X_{1,1}$, $\cdots$, X_{1,M_1} are called the factors of the social environment of P_1.

There are a number of other mathematical constructions with possible applications to social action theory, notably those of Rashevsky, Simon, Flood (4, 11, 14), which are different from the present one. The precise relations of the present concepts to those in the literature will not be described here. The decision as to which model is most fruitful depends, in part, upon the richness of the deductive structure which can be constructed with its concepts, the extent to which such theorems correspond to precise, nonobvious statements about social science which could be empirically verified, and the particular questions it is constructed to deal with. The present model is intended to deal with questions of stability and organizational efficiency in systems of action. It is fundamentally related to the theory of automata (22), a branch of applied mathematics in which significant results have been obtained, and in which there is currently a considerable amount of activity; certain theorems, concepts, and techniques in this field have direct applications for certain problems in action theory. In particular, if the behavior functions are taken as deterministic, e. g., Equations 1, which is tantamount to regarding the actor-unit as automata, the mathematical framework becomes similar to that used by Turing (25), Kleene (6), Moore (22), etc. If the probabilistic analogues of Equations 1, Expressions 2 are used, the conceptual structure is closely related to the automata with random elements studied by von Neumann (26), Murray (17), Shannon (22), etc. The models for action theory which arise from game and decision theory fail to take account of the dynamic aspects of action. There is, furthermore, no counterpart of the internal state variables, which permit the introduction of the subjective (goal) and normative aspects. Finally, the mathematical techniques of game theory have not yet been adequately extended to deal with "nonrational" behavior. Marschak's (14) decision-theoretic model assumes a deterministic behavior (the functions R_{I} and R_{II} in Reference 14, p. 192), also rational behavior, and does not consider internal state variables for storing information.

There are potent reasons for regarding the assumption of deterministic behavior functions as untenable. The following arguments support this view. Consider two actor-units, P_1 and P_2, and suppose that P_1 is a newborn infant, P_2 its parent. If P_1 and P_2 behaved automatically, that is, if the actor always responded unerringly in the same manner to the same situation, (e.g., the infant always cries when hungry, and the parent always gives him a milk bottle when he cries), then it is very unlikely that the infant ever would learn to use different signals for different needs, not to consider the acquisition of "knowledge" and "culture." Furthermore, such a mechanistic behavior must be a biological property of the actor-unit, which, since it is unchanged after birth, must be a purely inherited characteristic. This would be an inadequate assumption for a social actor-unit. In addition, an automatic type of behavior leaves no room for "voluntaristic" considerations. Finally, there is very little evidence of living organisms which do behave in a completely automatic manner.

To indicate how the proposed formal model leads to implications which may be significant for social action theory, the following results will be summarized in

nontechnical terms. They are rigorously stated and demonstrated elsewhere (9).

(*a*) If the deterministic assumption were retained, then for any initial situation, the behavior of the entire group will eventually be recurrently periodic, provided that the sets corresponding to all factors are finite. The particular nature of the asymptotic oscillatory behavior, and the approach to it, depend upon the types of behavior and the communication network of the actors.

(*b*) It is possible to define self-oriented and other types of deterministic behavior, and to show that such behavior types (strategies) exist under appropriate conditions.

(*c*) There is a systematic procedure for so selecting and combining actor-units with deterministic behavior as to produce any specified group behavior.

(*d*) If the probabilistic assumption is made, it is possible under certain conditions, to find an asymptotic distribution of the state of an actor-unit and its environment, which is independent of the initial states (17). The possibility of experimenting with such models on electronic computers as a guide to new conjectures, and disproof of false conjectures, is very promising as a research tool for social action theory. Numerical experiments of this nature have been made with varying degrees of success (5, 8, 18).

If the various concepts in the present model are suitably interpreted, the action schemes of Marschak, Parsons, and Levy can be obtained. It will be recalled that an "act" by actor P_1 is an element of the set X_1, which is specified to suit the problem, and which is interpreted as the collection of all possible observable changes the analyst wishes to attribute to P_1 during a small time period. Suppose that there are two natural factors in the natural environment, Y_1, Y_2. If the third equation of Equations 1 is of the form: $y_1^{h+1} = F(x_1^{h+1}, y_1^h, y_2^h)$, then it is possible for P_1's behavior function to be such as to make y_1^{h+1} take on any value in Y_1. In other words, P_1 can control the variable y_1^h, or the factor Y_1. Such a factor may be identified with the *means* corresponding to an act of P_1.[4] For instance, the location of an automobile at time $h + 1$ is a means of an actor relative to several acts involved in driving at time h. The values of the variables $x_1^h, \cdots, x_N^h, y_1^h, y_2^h$ specify the *situation* of P_1 at time h, in which the value of y_1^h specifies the state of his *means* and the remaining variables the *conditions* under which the act, x_1^{h+1} is selected. Suppose, for illustration, that P_1 has four factors which describe the internal state. Let $S_{1,1}$ represent the set of all "statements" or messages, stored in a suitably coded form, the content of which refers to the values of all factors at some future time period. Let $S_{1,2}$ represent the set of all degrees of "satisfaction" among which P_1 can discriminate, according to some hypothetical test. It is possible to construct the behavior function for P_1 so that

(*a*) $s_{1,2}^h$ indicates a high degree of "satisfaction" only if the values of all the factors at the present time period are "close" to $s_{1,1}^h$, and

(*b*) if they are not close, that x_1^{h+1} will be such as to approach closeness. The notion of "closeness" is a topological concept which has not yet, but must eventually be built into this conceptual framework. However, the notion of *goal-orientation* is thus seen to be a particular aspect of behavior. It is now also readily seen, how y_1^h, y_2^h become the "states of the world" in Marschak's one-actor model, and how the behavior function in the present model can be interpreted as his rule of action, R_1.

Consider a third factor of the internal state, $S_{1,3}$, which is interpreted as the set of all possible messages in coded form, the content of which refers to the values of the variables during some *past* time period. Any such message may be regarded as an observation by P_1. It is possible to define acts which merely signal the coded form of such a message to one or more other actors, thereby transmitting information. A behavior function which specifies when what

[4] In another paper, a distinction is made between means which depend on the actions of only one actor, and means which can only be "controlled" by a "cooperative" group of actors. This distinction will be avoided here.

message is to be sent to whom is simply a "rule of communication" in Marschak's model, and the "best" rules of action and communication are those which maximize $s^h_{1,2}$, the degree of "satisfaction." Finally, a fourth factor of the internal state of P_1, $S_{1,4}$, may be interpreted as the set of all possible rules or *norms* of behavior which have been learned by P_1 as the result of his interaction with $P_1, \cdots, P_N$ and E over the entire past. This suggests very strongly that:

(*a*) the behavior function of P_1 be probabilistic, i.e., of the form of Expression 2 rather than deterministic, and

(*b*) that the stochastic process (Markoff) so defined be nonstationary; i.e., that the conditional probability given by Expression 2 be permitted to vary with time. (In future models it may be profitable to explore more general stochastic processes, where the dependence extends more than one step into the past.) Thus, the present method of formalizing an action scheme, includes the basic concepts of Parsons, as well as those of Marschak and the approaches they typify. It also meets the requirements of Levy's definition insofar as an act, x_i^{h+1}, is determined by factors which do not refer to environmental and/or hereditary aspects of the species only.

II. A SUBSTANTIVE DEFINITION OF ACTION

The concept of action just formalized is "residual" in the sense that it presents a mathematical framework, the elements of which still require sociological interpretation. That is, the number and kind of basic variables, and certain relations between them have yet to be axiomatically specified such that they provide:

(*a*) a logically consistent and precise mathematical system, which is also interesting (challenges efforts of mathematicians), fruitful (rich in potential conjectures and theorems), and such that

(*b*) its symbols and formal relations (which are devoid of any intrinsic empirical content) can be so interpreted in sociological contexts as to reflect some empirical "truths," and also readily admit the possibility of widening the scope of empirical situations to which applications can be made.

Such a substantive specification of the basic elements of a starting model for the social action concept is proposed here within the conceptual framework proposed in Part I. This effort appears to be completely unrelated to the work of Zelditch (30), who has worked on the same sort of problem. To our knowledge the approach to "Learning" which is involved here has not been treated in the literature. It is ultimately desired to explore the implications of certain "plausible" assumptions, based on some of the available evidence in psychology, rather than introduce parameters which must be empirically estimated, and this study is a first step in that direction. The relation to these other approaches is, however, readily obtained.

It is intended to formulate some of the basic concepts with some precision in order to lay the groundwork for further research, both along empirical and theoretical lines. No claims to the rigor requisite for thorough mathematical analysis are made, and it is not believed that a complete mathematical model should be constructed before more detailed assumptions, supported by empirical data, are added. The assumptions and concepts proposed here, however, are believed to have more than a tentative significance, and to permit the construction of experiments and the formulation of important problems in social action theory.

Informal summary of procedure

Nonmathematically trained readers may see the relevant part of the summarizing discussion at the end of this paper for a completely verbal statement of the core of what follows.

To emphasize the main concepts, without losing generality, it is desirable to confine the discussion to a single individual, denoted by P, and his environment, $\bar{P}$, relative to a specified universe of discourse. No properties other than those specifically assigned to P are to be adduced to P. (In particular, it is not to be inferred that P is a normal human adult.) The first property to be ascribed to

P is that at any instant of time h, P is to be described by the values of four variables s_1^h, s_2^h, s_3^h, x^h and $\bar{P}$ by the variable $\bar{x}^h$. These variables range over the sets S_1 (assumed finite), S_2 (assumed infinite), S_3 (assumed infinite), X^h (assumed finite and changing with h) and $\bar{X}^h$ (assumed finite). The basic questions regarding P refer to the manner in which s_1^{h+1}, s_2^{h+1}, s_3^{h+1} and x^{h+1} are related to s_1^h, s_2^h, s_3^h, x^h, $\bar{x}^h$. In Part I it was argued that such a relationship ought to be probabilistic. In this section, these symbols will be given a definite interpretation, and some reasons for the choice of these particular interpretations and the number of variables will be advanced. Then, some assumptions about the proposed relations will be informally discussed.

The set S_1 is called the "biological" factor, and consists of all possible biological states in which P can be observed to be at any instant. s_1^h is such a state at epoch h. Inasmuch as the complete specification of S_1 is regarded as a task for a biologist and hence is not undertaken here, the subsequent definition of action based on this is still "residual," but not to the extent it was previously. A preliminary list for the elements of S_1, with P interpreted as a neonate, could be the following, $b_1 \equiv$ "being cold," $b_2 \equiv$ "being hungry," $b_3 \equiv$ "being dead," $b_4 \equiv$ "being in pain," $b_5 \equiv$ "none of these," $b_6 \equiv$ "b_1 and b_2," $b_7 \equiv$ "b_1 and b_3" etc.

$\bar{P}$ will be interpreted as a parent, who has as one of his objectives to "gratify" P as efficiently as possible. Suppose that $\bar{X}$ is the set of all possible actions $\bar{P}$ could choose during a specified time interval, say 1 second. For instance, $\bar{X}$ may be a list such as: $\bar{a}_1 \equiv$ "making a sound," $\bar{a}_2 \equiv$ "part of a motion in presenting food, as a milk bottle," $\bar{a}_3 \equiv$ part of a motion in applying warmth, as a blanket," $\bar{a}_4 \equiv$ "part of a motion in changing diapers," $\bar{a}_5 \equiv$ "doing nothing," $\bar{a}_6 \equiv$ "$\bar{a}_1$ and $\bar{a}_2$," $\bar{a}_7 \equiv$ "$\bar{a}_2$ and $\bar{a}_3$" etc.

Let Ξ be the list of all possible actions which P is capable of performing during, say, one second, at any time in its "life." Associate a certain set of subsets of Ξ to every instant h, (a partition of Ξ), and call the resulting set X^h. An element of X^h, x^h, is a general heading for one or more items in the list Ξ. Thus, for $h = 0$ (the instant of birth), it might be reasonably supposed that X^h consists of the following elements: $a_1 \equiv$ "making a sound," $a_2 \equiv$ "making some motion," $a_3 \equiv$ "emitting an odor," $a_4 \equiv$ "doing nothing," etc.

The variables $\bar{x}^h$, s_1^h, and x^h are called the input, biological state, and output, of P at time h, respectively. It may happen that, whenever P is in the biological state s_1^h, there is an input sequence of some length, H, $\bar{x}^h$, $\bar{x}^{h+1}$, $\cdots$, $\bar{x}^{h+H}$, which $\bar{P}$ might present to P, such that P responds with the output sequence x^{h+1}, x^{h+2}, $\cdots$, x^{h+H+1}, but that such an input sequence produces the same output sequence for no other s_1^h. Such an input sequence may be regarded as an experiment by $\bar{P}$ on P, with the output sequence as observed result, and information about s_1^h as the inferred conclusion. It is assumed that for every s_1^h, there is some such experiment (an input sequence associated with a unique output sequence), and that no single experiment refers to two different states s_1^h. States for which this does not hold are excluded from S_1. Two states are called indistinguishable if the same output sequence results when the associated input sequence is applied to P in each of the two states. Two individuals, P and P', are called biologically similar if for every biological state of P there is exactly one biological state of P' from which it is indistinguishable, and vice versa. It is felt that the factor S_1 is essential in any substantive specification for the social actor, since it is desired to aim this development toward actor-units which are biological organisms, in particular, human beings.

The set (or factor) S_3 is called the evaluative factor. Its elements are the various possible gradations of preference or satisfaction which are to be determined on the basis of a questionnaire, or preference test of some sort. They are the values which the variable a_3^h, called an indicator, can take at any h. It is reasonable to suppose that S_3 consists of at least three items: $B =$ "bad," $I =$ "indifferent," and $G =$ "good." It is assumed that some element of S_3 is associated with every pair of elements from S_1 and

TABLE 1

h =	0	1	2	3	4	5	6	7	8	9	10
s_h =	b_2	b_2	b_2	b_2	b_2	b_2	b_2	b_5	b_5	b_5	b_1
s_3 =	—	B	B	B	B	B	B	G	G	G	G
x^h =	a_1	a_1	a_1	a_1	a_4	a_4	a_4	a_4	a_4	a_4	a_4
$\bar{x}^h$ =	$\bar{a}_1$	$\bar{a}_3$	$\bar{a}_4$	$\bar{a}_2$	$\bar{a}_2$	$\bar{a}_2$	$\bar{a}_2$	$\bar{a}_2$	$\bar{a}_5$	$\bar{a}_5$	$\bar{a}_5$

etc. Note that there is a lag (h = 4–7) between stopping to cry and feeling satisfied.

$\bar{X}$; that is, a "value judgment", S_3^h, is "experienced" by P at time h for any combination of the biological state, s_1^h and the input $\bar{x}^h$ of P at time h.

Closely related to S_3 and S_1 is the "cognitive" factor, S_2. In order to take into account the fact that the probability of s_3^h being B decreases with h to a minimum, without a teleological assumption, such as the tendency of P to maintain itself, it is supposed that

(*a*) P can acquire and retain information about the relations between s_1^h, x^h and s_3^h.

(*b*) P can respond to s_1^h with x^h so as to "maximize" s_3^h.

(*c*) The set of responses available to P, X^h, increases with h.

The mechanism which is postulated to enable P to execute *a* is the following: P is regarded to be endowed with a storage mechanism of great capacity, which can store information in coded form. The precise mechanism is yet unknown; it might conceivably consist of 10^{10} neurons, each of which can exist in one of two stable states, thus storing for an instant up to 10^{10} bits of information in discrete form, like a digital computer; or, it might store it in some other, e.g., analogue, form. The precise neurophysiological details of the storage mechanism are probably not essential to the following discussion. Thus, assuming a discrete storage mechanism after McCulloch and Pitts (15), all possible statements, encoded into a binary alphabet (e.g., Morse code without spaces) which could be so stored (or formed in such a storage) and "read out", are considered. If the content of each such statement which is meaningful to P, refers to an item of information about the relation between s_1^h, x^h and s_3^h, it qualifies as an element of S_2. At any time h, s_2^h describes the state of the storage element of P.

It is now possible to discuss some proposed relations between the variables. First, it can be supposed that, for small values of h, (in the early infancy of P), the response x^h and s_1^{h+1} are deterministically (to a good first approximation, with a sufficient number of biological and environmental factors considered) related to s_1^h and $\bar{x}^h$. At any such time, h, P is in biological state s_1^h (e.g., hunger), and automatically responds with x^h. Then $\bar{P}$ responds to x^h with $\bar{x}^h$, but not necessarily automatically. Consider now as a second supposition that the duration between h and $h + 1$ be sufficiently small so that $s_1^h = s_1^{h+1}$ with high probability for many values of s_1^h. Next, it is supposed that some biological mechanism assigns a value of s_3^h to the pair s_1^h, $\bar{x}^h$, and automatically evokes the response x^h (e.g., crying) and the new biological state s_1^{h+1} (e.g., still hungry). Table 1 indicates such a scheme. (For definition of symbols, see page 273).

A complete description of this type of early behavior for P involves a table which specifies x^h and s_1^{h+1} for every possible pair $\bar{x}^h$, s_1^h, or a functional representation. One possible specific form for the functional description of a biopsychological mechanism is:

$x^h = a_4$ if $s_1^h \bar{x}^h = b_1\bar{a}_3$, $b_2\bar{a}_2$, $b_5\bar{a}_5$, $b_6\bar{a}_7$, etc.

$x^h = a_1$ otherwise, with high probability.

$s_1^{h+1} = b_5$ if $s_1^h \bar{x}^h$, etc., with high probability.

$s_1^{h+1} = s_1^h$ otherwise.

Verbally the equations say:

The baby does not cry if: when it is cold, it is covered with a blanket; when it is hungry, it is fed; when it is comfortable it is left alone; when it is cold and hungry it is covered and fed, etc.

The baby cries otherwise.

The baby is comfortable if (etc. as before).

The baby continues in the same state otherwise.

There is, of course, no possibility for P to

learn how to maximize s_3 with this paradigm; that is, P cannot improve his chances of obtaining the appropriate $\bar{x}^h$ for any s_1^h by means of a suitable choice of x^h. Furthermore, $\bar{P}$ cannot improve his chances to choose the "correct" $\bar{x}^h$ (the one appropriate to s_1^h) on the basis of x^h with these crude assumptions. In order for P to "learn," information must be acquired. It could be assumed that P is naturally endowed with a set of responses such that exactly one response corresponds to every biological state. In this case, it is only the parent who needs to learn his appropriate response to a signal. An alternative and possibly more realistic assumption is that whenever an event like s_1^h, x^h, s_3^h, occurs, it is the "subjective feeling" accompanying it which is stored, (i.e., remembered, perceived). The "subjective feeling" is an entity like s_3^h, whereas s_3^h indicates the value judgment about $(s_1^h, \bar{x}^h)$, the former indicates the type of feeling (not describable in meaningful symbols) which corresponds to s_1^h for P. It is, of course, a nonobserved entity, and will attain meaning through the relations into which it enters, and which will lead to operationally meaningful entities. To discuss it, it will be convenient to abbreviate the phrase "subjective feeling of P corresponding to s_1^h, x^h, s_3^h, experienced at time h" by the symbols σ_1^h, ξ^h, σ_3^h. This concept will prove useful also in connection with P's subjective perception of x^h, $\bar{x}^h$, s_2^h and s_3^h. Denote the subjective interpretation of these by P with ξ^h, $\bar{\xi}^h$, σ_2^h and σ_3^h respectively. σ_1^h is, in a sense, that which is invariant under translation from one form of expression to another.

In general, it is also possible to consider the subjective interpretations of x^h, $\bar{x}^h$, $\bar{s}_2^h$, and $\bar{s}_3^h$ by $\bar{P}$, and these may be denoted by θ^h, $\bar{\theta}^h$, $\bar{\sigma}_2^h$, $\bar{\sigma}_3^h$, respectively. Correspondingly, the following conditional probabilities would also be specified: $\prod_x(\xi^h \mid x^h)$, $\prod_{\bar{x}}(\bar{\xi}^h \mid \bar{x}^h)$, $\prod_x(\theta^h \mid x^h)$, $\prod_{\bar{x}}(\bar{\theta}^h \mid \bar{x}^h)$ and $\prod_\sigma(\sigma_1^{h+1} \mid s_1^h)$. This admits the possibility of "misinterpretations" and errors. Since this is regarded as a second stage in the development of this model, it will be postponed for a future project, in which the problem of noise and error will be treated more generally. For the present, it shall be assumed that the sets of all ξ^h, $\bar{\xi}^h$, σ_1^h, σ_2^h, σ_3^h and x^h, $\bar{x}^h$, s_1^h, s_2^h, s_3^h are in 1–1 correspondence, and that the sets of x^h, $\bar{x}^h$, $\bar{s}_2^h$, $\bar{s}_3^h$ and θ^h, $\bar{\theta}^h$, $\bar{\sigma}_2^h$, $\bar{\sigma}_3^h$ are also in 1–1 correspondence. Henceforth, only the variables x^h, $\bar{x}^h$, s_1^h, s_2^h, $\bar{s}_2^h$, s_3^h, $\bar{s}_3^h$ shall be used and the distinction between the operationally defined quantities and the subjectively perceived versions of these shall be temporarily disregarded.

Under this assumption, s_2^h, the item of information stored at time h, consists of s_1^h, x^h, s_3^{h+1}. After both P and $\bar{P}$ have learned say at time k, P can, on the basis of s_1^k, select a response, x^k, which will be very likely to evoke the appropriate response from $\bar{P}$, $\bar{x}^k$. What mechanism could be postulated to account for such learning? A concrete, hypothetical example may illustrate the point. P is in state $s_1^0 = b_2$ (hungry) and responds with $x^0 = a_1$ (cries), has $s_2^0 = 0$ (knows nothing) $s_3^0 = B$ (is dissatisfied); $\bar{P}$ responds with $\bar{x}^0 = \bar{a}_3$ (makes sound). Next, P is in state $s_1^1 = b_2$ (still hungry), $x^1 = a_1$ (still cries), $s_2^1 =$ (x^0 (its own impression of having cried), $\bar{x}^0$ (its impression of $\bar{P}$'s sound), s_3^0, s_1^0 (its impression of having been hungry)), $s_3^1 =$ (still dissatisfied); $\bar{P}$ responds with $\bar{x}^1 = \bar{s}_2$, etc. The subjective perception of hunger is then associated with a milk bottle as leading to satisfaction, and having been evoked by crying. This item of information is stored. Each time it recurs, a tally is added to the previous count in the storage of this item. Later, other responses than crying are tried by P when hungry, and tallies are distributed among those responses. In a given state s_1^h, P selects that response with the largest number of tallies with high probability, or tries a new response at random, so as to achieve satisfaction most rapidly.

It is suggestive to study the conditional probability that P selects response x^h at time h given that the state was s_1^h. It will be of interest to determine how this probability changes with time, and what sort of asymptotic behavior may result. It may immediately be guessed that the probability at any time is proportional to the number of

tallies associated with the response, with "new" responses being equiprobable with a probability not exceeding that of any response for which there is one tally. It should be noted that the response, x^h, permits P to acquire more information in the sense of storing more items of experience, and also to satisfy its needs in the sense of maximizing s_3^h. P is said to learn, when his behavior changes systematically, i.e., if the uncertainty of P about which response to choose next for a given $s_1^h \bar{x}^h$ decreases, or equivalently if the average number of responses required to obtain gratification decreases.

Another approach is based upon the assumption that the storage mechanism of P and $\bar{P}$ is of an "analogue" nature, that time varies continuously, and that s_2^h represents a real function of several integral variables with the following interpretation: Let t denote the time in place of h, and suppose that a real number, s_2, is assigned to every real t and every triplet of integers x, s_1, and s_3. That is

$s_2 = s_2(x, s_1, s_3, t)$, and it is interpreted as the "strength" or "propensity" of P toward the response x at time t, at which P is in the biological state s_1, with the resulting value s_3. It is assumed that s_3^t is determined by s_1^t and $\bar{x}^t$. To simplify the concepts, suppose further, that for every s_1^t, there is exactly one pair $(x^t, \bar{x}^t)$ which makes $s_3^t = G$ (i.e., is rewarding). Suppose, further, that $\bar{x}^t$ is a unique function of x^t, so that for every biological state of P at time t, there is a unique response, x^t, which will result in reward to P. Then $s_2(s_1, t)$ denotes the strength of the "correct" response to the state s_1^t. The advantage of this simplified formulation is that a well-known model of learning may be applied directly.

Precise formulation of learning concepts

Let $p^h(x^h \mid s_1^h)$ be the probability defined above. Suppose that the set X has n elements, $a_1, \cdots, a_n$. Because p^h is a probability, it is necessary that

$$\sum_{x^h=a_1}^{a_n} p^h(x^h \mid s_1^h) = 1 \text{ for every } s_1^h \text{ and all } h. \quad [3]$$

As mentioned previously, an item of information which is stored by P at time h consists of a triplet of values of the variables s_1^h, x^h, s_3^h. This triplet, translated into the code in which information is stored, is one of the components which specify s_2^h. Another component is $n(s_1^h, x^h, s_3^h)$, the number of times that this triplet has occurred up to time h. Now let

$$\gamma(s_1^h, x^h, s_3^h) = \frac{n(s_1^h, x^h, s_3^h)}{\sum_{x^h}\sum_{s_3^h} n(s_1^h, x^h, s_3^h) + c(s_1^h)} \quad [4]$$

Let $\tilde{n}(s_1^h)$ be the number of x^h such that $n(s_1^h, x^h, s_3^h) = 0$ for *any* s_3^h, and let $c(s_1^h) = n - \tilde{n}(s_1^h)$. Observe that

$$1 = \sum_{x^h}\sum_{s_3^h} \{\gamma(s_1^h, x^h, s_3^h) + c(s_1^h)/[\sum_{x^h}\sum_{s_3^h} n(s_1^h, x^h, s_3^h) + c(s_1^h)]\}.$$

The main assumption stated previously, can now be written as:

$$\left.\begin{aligned} p(x^h, s_1^h) &= \sum_{s_3^h} \varphi(s_3^h)\gamma(s_1^h, x^h, s_3^h) \quad \text{if } \tilde{n}(s_1^h) \neq 0 \\ p(x^h, s_1^h) &= 1/c(s_1^h) \text{ if } \tilde{n}(s_1^h) = 0 \end{aligned}\right\} \quad [5]$$

where $\varphi(s_3^h)$ are such as to satisfy:

(*a*) Equation 3 for any γ.

(*b*) $0 \leqq p(x^{h+1} \mid s_1^h) \leqq 1$ for every s_1^h, and x^{h+1} and all h.

(*c*) $\varphi(s_3^h)$ is a nondecreasing function of s_3^h.

The following lemmas follow immediately:

(*a*) $0 \leqq \gamma(s_1^h, x^h, s_3^h) \leqq 1$ for all values of the variables.

(*b*) $$\sum_{s_3^h} \varphi(s_3^h)\sum_{x^h} \gamma(s_1^h, x^h, s_3^h) + \frac{\tilde{n}(s_1^h)}{c(s_1^h)} = 1 \text{ for all } s_1^h.$$

(*c*) If $\gamma(s_1^h, x^h, s_3^h)$ and s_3^h are large for a given s_1^h, x^h, then $p(x^h \mid s_1^h)$ is large.

$\varphi(s_3^h)$ represents a weighting function which indicates how the value judgments of P about the experienced states affect the responses of P to a given s_1^h, at present. Equation 5 is interpreted as determining the response of P based upon the frequency with which this response has resulted in a certain value (e.g., reward) in the past, weighted by the value; all the unprecedented responses, however, also have a certain

small probability of being selected. $\varphi(s_3^h)$ may be taken as a mapping of the set S_3 into the interval $-1 \leqq 0 \leqq 1$. For example, the value G may map into .9, I (indifferent) into 0, and B into $-.8$. Thus $p(x^h \mid s_1^h)$ would be a linear combination, like .9 $\gamma(s_1^h, x^h, s_3^h = G) - .8\ \gamma(s_1^h, x^h, s_3^h = B)$.

Let $q(s_1^{h+1} \mid s_1^h, \bar{x}^h)$ be a conditional probability, which must be specified. It is such that for most pairs s_1^h, $\bar{x}^h$, $q(s_1^{h+1} = s_1^h \mid s_1^h, \bar{x}^h) \cong 1$. There are certain pairs, s_1^h, $\bar{x}^h$, which will be called "appropriate," for which $q(s_1^{h+1} \neq s_1^h \mid s_1^h, \bar{x}^h)$ is small, For such pairs, it is assumed that $s_3^h = V(s_1^h, \bar{x}^h) = G$. If there are m elements in the set S_1, it may be supposed that $q(s_1^{h+1} \mid s_1^h, \bar{x}^h) = 1/m - 1$ if $V(s_1^h, \bar{x}^h) = G$ for any $s_1^{h+1} = s_1^h$. Furthermore, it is assumed that:

$$Pr(x^h, s_1^{h+1} \mid \bar{x}^h, s_1^h) = p(x^h \mid s_1^h) \cdot q(s_1^{h+1} \mid s_1^h, \bar{x}^h). \qquad [6]$$

Next, let

$$r(\bar{x}^h \mid x^h) = \sum_{\bar{s}_3^h} \bar{\varphi}(s_3^h) \cdot \bar{\gamma}(x^h, \bar{x}^h, \bar{s}_3^h) \qquad [7]$$

where

$$\bar{\gamma}(x^h, \bar{x}^h, \bar{s}_3^h) = \frac{\bar{n}(x^h, \bar{x}^h, \bar{s}_3^h)}{\sum_{\bar{s}_3^h} \sum_{\bar{x}^h} \bar{n}(x^h, \bar{x}^h, \bar{s}_3^h)} \qquad [8]$$

Equation 7 is similar to Equation 5, except that the biological state for $\bar{P}$ and the possibility of choosing new, unprecedented responses are ignored for reasons of simplicity. Since $\bar{P}$ is interpreted as a parent, and P as an infant, it may be assumed that the value scale of $\bar{P}$ is such that the highest value, $\bar{s}_{3\max}$, coincides with the highest value for P after both P and $\bar{P}$ have learned. The only information available to $\bar{P}$ is x^h, and $\bar{s}_3^h$ is therefore a function of x^h. Thus, any x^h for which $\bar{s}_3^h = \bar{V}(x^h) = \bar{s}_{3\min}^h$ is said to be interpreted by $\bar{P}$ as an act of rejection, and $\bar{P}$, striving to evoke a response x^h from P such that $\bar{V}(x^h) = \bar{s}_{3\max}^h$, will modify his output. If an act by P, x^h, which was interpreted by $\bar{P}$ as rejection, was, in fact, a consequence of the pair s_1^h, $\bar{x}^h$ for which $V(s_1^h, \bar{x}^h) = s_{3\min}^h = B$, the change $\bar{x}^{h+1} \neq \bar{x}^h$ can increase s_1^h with greater probability than $\bar{x}^{h+1} = \bar{x}^h$. It is possible for P to learn this fact. If, then, P continues to respond with x^h (which is interpreted as rejection of $\bar{x}^h$) to the state s_1^h the average time for s_3 to become G is less than if P responded at random. P, in such a case, is said to have acquired "command over the generalized negative."

In such a case, P has learned that one particular response (the one which is interpreted by $\bar{P}$ as rejection, e.g., crying) is effective in achieving satisfaction. Furthermore, P will have acquired information about which $\bar{x}^h$ are inappropriate or unsatisfying with what s_1^h. $\bar{P}$ will not have acquired this information since $\bar{P}$ does not store s_1^h. $\bar{P}$ will not learn which $\bar{x}^h$ should correspond to what x^h, to maximize $\bar{s}_3^h$, or what is equivalent, s_3^h, and P will not learn which x^h should correspond to what s_1^h in order to maximize s_3^h (by evoking the appropriate $\bar{x}^h$).

It is clear that $\sum_{s_3^h} \sum_{s_1^h} n(s_1^h, x^h, s_3^h) = \sum_{\bar{x}^h} \sum_{s_3^h} \bar{n}(x^h, \bar{x}^h, \bar{s}_3^h)$, which is the number of times up to time h that the response x^h, has occurred. This number can be experimentally determined.

It may be remarked that there are two types of variables involved in this approach. The first kind, consisting of x^h, $\bar{x}^h$, s_1^h have definite empirical referents, which have both an objective and a subjective significance. These variables appear in the three basic functions

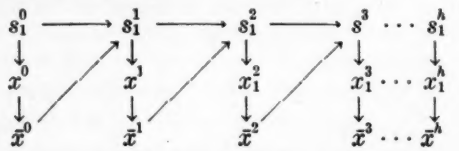
$$p(x^h \mid s_1^h), \qquad q(s_1^{h+1} \mid s_1^h, \bar{x}^h), \qquad r(\bar{x}^h \mid x^h).$$

The second kind of variables, consisting of s_3^h, $\bar{s}_3^h$, and the functions in which they appear, $n(x^h, s_1^h, s_3^h)$ and $\bar{n}(x^h, \bar{x}^h, \bar{s}_3^h)$, are auxiliary variables. They enter into the formulation of assumptions about the form of the basic functions, but do not appear in any final result. The functions n and $\bar{n}$ may be regarded as one form of the cognitive variables s_2^h and $\bar{s}_2^h$. The interpretation of s_2 as a "strength" is a different form of this variable, and the theoretical development of this concept will follow in the summarizing discussion.

There are two types of problems corresponding to these types of variables. First, to determine what theorems can be

established without assuming special forms for the basic functions, and their significance for action theory. The auxiliary functions play no role here. Second, to explore the consequences of various assumptions which restrict the form that the basic function can have, and determine their relation to empirical data. These assumptions are first formulated in terms of the auxiliary functions $n(s_1^h, x^h, s_3^h)$ and $\bar{n}(x^h, \bar{x}^h, \bar{s}_3^h)$, and later, in terms of the strength s_2^t. It is a matter of some interest to determine how wide a range of varying assumptions of this nature still yield the same fundamental theorems.

With regard to the first type of problem, the time sequence of the basic "perceived" variables may be diagrammatically represented as follows:



An arrow terminating at a variable means that this variable "depends" (in the sense of conditional probability) on the variable from which the arrow originates. The probability of such a time sequence is given by $e(s_1^0, \cdots, s_1^h; x^0, \cdots, x^h; \bar{x}^0, \cdots, \bar{x}^h) = p(x^0/s_1^0) \cdot q(s_1^1/\bar{x}^0, s_1^0) r(\bar{x}^0/x^0) \cdot p(x^1/s_1^1) \cdot q(s_1^2/\bar{x}^1, s_1^1) \cdot r(\bar{x}^1/x^1) \cdot p(x^2/\bar{x}^2, s_1^2) \cdot q(s_1^3/\bar{x}^2, s_1^2) r(\bar{x}^3/x^2) \cdot \cdots \cdot p(x^h/\bar{x}^h, s_1^h) \cdot r(\bar{x}^h/x^h)$. The probability of the triplet $x^h, s_1^h, \bar{x}^h$ given $s_1^0 x^0 \bar{x}^0$ is given by

$$\sum_{s_1^1 \cdots s_1^{h-1}; x^1, \cdots, x^{h-1}; \bar{x}^1, \cdots \bar{x}^{h-1}} e(s_1^0, \cdots, s_1^h; x^0, \cdots, x^h; \bar{x}^0, \cdots \bar{x}^h) \equiv E(x^h, s_1^h, \bar{x}^h \mid s_1^0 x^0 \bar{x}^0).$$

If $\sum_{s_1^h} \sum_{x^h} \sum_{\bar{x}^h} E(s_1^h, x^h, \bar{x}^h \mid s_1^0 x^0 \bar{x}^0) = 1$ for every triplet $s_1^0 x^0 \bar{x}^0$, then it can be shown by matrix algebra, that, as $h \to \infty$, E does not depend on $s_1^0, x^0, \bar{x}^0$ (17). This condition requires that there be no terminal states, that is, triplets $s_1^h, x^h, \bar{x}^h$ for which the probability of change is zero. The above result is typical of the type of theorems which can be obtained for the first type of problem.

With regard to the second kind of problem, using the auxiliary functions n and $\bar{n}$, it is convenient to investigate the problem at $h = 0$ in the following phases: (*a*) Suppose that s_1^0 occurs at random, that is, if there are m biological states, suppose that the probability of any one occurring at $h = 0$ is $1/m$. (*b*) Since P has not yet stored any relevant information and since $\bar{P}$ has not yet stored any additional information relevant to this specific relation it may be supposed that any one of the n possible responses of which P is biologically capable occurs with probability $1/n$ (or any other a priori distribution). (*c*) Similarly, suppose that any of the $\bar{n}$ responses of $\bar{P}$ occurs with probability $1/\bar{n}$. (*d*) s_3^0 is determined according to a known function, $V(\bar{x}^h, s_1^h)$. (*e*) P stores the item of information: "s_1^0, x^0, s_3^0," which is the value of the variable s_2^0. Suppose that for each s_1^h, there is only one $\bar{x}^h$ such that $V(\bar{x}^h, s_1^h) = s_3^h = G$, and for each $\bar{x}^h$ only one s_1^h such that $V(\bar{x}^h, s_1^h) = G$. Then $Pr\{s_3^0 = G \mid s_1^0\} = 1/n$. Thus, $Pr\{s_1^1 = s_1^0\} = (n-1)/\bar{n}$, and $Pr\{s_1^1 \neq s_1^0\} = 1/\bar{n} \cdot 1/m$ for each of the $m - 1$ states not equal to s_1^0. The cycle now repeats. If $s_3^0 = B$, P will have remembered that x^0 in association with s_1^0 did not result in reward, hence should not be selected at $h = 1$, since $s_1^1 = s_1^0$. The probability of x^1 can therefore be taken as $1/(n-1)$. At this point the value to $\bar{P}$ is determined on the basis of x^1, but it is the value corresponding to $\bar{x}^0$, so that it will be denoted as $\bar{s}_3^0$ rather than $\bar{s}_3^1$. Certain responses of P make $\bar{s}_3^0 = G$, others $\bar{s}_3^0 = B$. It may be supposed that the responses x^1 which make $\bar{s}_3^0 = G$, are also associated with a state, s_1^1, for which any response, $\bar{x}^1$ will make $s_3^1 = G$. Now, $\bar{x}^1$ is selected by $\bar{P}$, and, since the situation for $\bar{P}$, x^1, is probably not the same as at $b = 0$, the selection is made at random.

This process continues, and it is clear that soon certain combinations of $s_1^h, x^h, s_3^h = G$ will have occurred more frequently than others, and the probability of these x^h in response to the s_1^h in such combinations will be larger than that of other responses. Thus, the frequency of such combinations will increase still further until $p(s^h/s_1^h) \to 1$ for every s_1^h. In other words, $x^h = f(s_1^h)$ when

P has learned. Similarly, certain combinations of x^h, $\bar{x}^h$, $\bar{s}_3^h = G$ will occur more frequently than other combinations, and $r(\bar{x}^h/x^h) \rightarrow 1$, or $\bar{x}^h = \bar{f}(x^h)$ when $\bar{P}$ has learned. This argument demonstrates that, under the above assumptions on the basic functions, learning by P and $\bar{P}$ will always take place, and when it is completed, s_3^h and $\bar{s}_3^h$ will be high for all h. This result, plus the following additional reasons support the view that Equations 5 and 7 are tenable assumptions:

(*a*) These assumptions represent a simple theoretical scheme permitting learning to occur. They refer to the general properties of any storage device without regard to the substantive content.

(*b*) They are intuitively plausible.

(*c*) There are certain points of overlap with other learning models, and the present formulation is in a certain sense simpler.

As h increases, the amount and kind of information acquired become very complex, and two distinct problems emerge: How much, i.e., what, is learned, and how well is it learned. Even so, these concepts are not sufficient to account for but one small aspect of human learning. It is expected, however, that the research in this direction will lead toward a fruitful means of describing social action, but before further steps can be taken, it is felt necessary to isolate and analyze the logical and mathematical structure of this model much further and more precisely than has been attempted here.

III. THE LOGICAL AND RATIONAL ASPECTS OF ACTION

Man is frequently called a rational animal. This description refers to a classification of animal behavior into rational and nonrational. Such classifications have often been proposed, and many of them are related to Pareto's reflections on this topic (19). In his view, an action is called logical if the subjective ends are identical with the objective ends, i.e., if the means are suited to the ends. In recent years, the developments in mathematical economics (14), game theory (27), and decision theory (28) have led to various concepts of rationality. These usually involve the assumed assignment of computational facilities of considerable complexity to the actor, by means of which he might select one from a specified set of alternative choices. To meet the requirements for rationality, such an actor must select that choice which optimizes his personal utility or satisfaction, according to an arbitrary criterion (minimax, Bayes, Laplace, Hurwitz).

In Parts I and II of the present paper, the foundations of a formal conceptual scheme for action theory, as it has been developed by Pareto, Durkheim, Weber, Parsons, Levy have been laid in a preliminary manner. The first of these attempts (Part I), in which the residual nature of current definitions of the concept of social action was discussed with some precision, was designed to clear the ground for the development attempted in Parts II and III. Part II dealt with the acquisition of cognition by the neonate actor, and involved the precise formulation of general hypotheses abou certain aspects of social learning. Cognition so acquired emerges as an element in, or aspect of, all action which cannot be adequately accounted for in terms of the factors of species heredity and the nonspecies environment.

In summary, an action scheme was considered to involve several actor-units and their "natural" environment. The actors are specified by internal state variables, output (action) and input (actions of others and of environment) variables, which are defined for each of a discrete number of time instants, and the following relations between them: the joint conditional probability of the output and the internal state at any instant, given the input and the state at the previous instant is assumed to be given for each actor and is called his behavior function. Now it is proposed to study the classification of such behavior functions into logical-nonlogical and rational-nonrational categories, and to compare this classification with the related efforts in the literature.

It should be emphasized that many of the terms which may be used here, such as rational, logical, cognitive, etc., do not have a meaning which is uniformly accepted or consistently used. The concepts they denote

shall here be defined by their formal logical properties. In this section the concepts shall be referred to by their names rather than by symbols, and although these names may have misleading connotations, it should be remembered that what they denote will be formally defined by a set of properties, nothing more. Thus, the term "cognitive" shall refer to those aspects of an actor-unit which deal with the transfer, transformation, and storage of information, i.e., to the internal state-variables and the manner in which they change. To the extent that this approach proves fruitful, it would argue that the usual distinctions among cognitive orientations, cathectic orientations, and evaluative orientations (20) are not on the same sort of level of generalization ordinarily proposed but rather represent different states of cognition in this more general sense. With recent developments, both mathematical and empirical, in information theory and allied fields, the approach to general analysis of action via cognition in the sense of the term used here has certain prospects of fruitfulness. This usage avoids the major pitfall involved in the frequent past implicit or explicit emphasis in the role of cognition in action since it does not necessarily impute to any given instance of action either conscious understanding on the one hand, or on the other, logicality or rationality as the terms are subsequently defined in this paper.

Types of ends. An end is regarded as being specified by the values of certain internal state variables at some instant of time. The values of these variables are the characters (symbols) of a given code (e.g., dots and dashes of Morse code, "bits" of computer codes, letters of our alphabet, firing or nonfiring of nerve cells, etc.). The combination of these values at a given time spells out a statement or message, the content of which refers to a future "anticipated state of affairs." It is assumed that this content is "meaningful" to the actor-unit. It will be convenient to refer to such internal state variables as end-variables. In addition, it is required that a certain other internal state variable, called an indicator, indicates at every instant to what extent the actual state of affairs is "close" to the "anticipated state of affairs," as encoded in terms of the end-variables. Such an indicator may, in special cases, be interpreted as the degree of utility or satisfaction associated with a state of affairs for the actor in question, because it establishes a "preference" ordering (partial) on the set of all such states. The "future state of affairs" to which the statement which is coded in terms of the end-variables refer, is of two kinds:

(*a*) It may be specified by the values of certain variables for the "natural" environment or for the actions of others at a future time period (possible several instants). In this case the ends are called *objective* ends.

(*b*) It may be specified by the values of certain internal state variables of the actor at a future time period. In this case the actions are called *nonobjective*. Cases where it is specified by the values of variables some of which refer to the internal state and some of which to the actions of others and the "natural" environment are readily visualized. In such cases the ends are called *mixed*, or are not classified at all.

Types of means. It is possible to consider actions and "natural" variables which in addition to objective (i.e., empirical) consequences as magic rituals, supplications, etc. are primarily significant from the point of view of the actor for other reasons (12). A classification of the means into objective and nonobjective is thus suggested. It should be remembered that a means in the present context is a variable of the "natural" environment which depends upon the "output" or "action" of an actor. In turn, the variables which specify certain future states of affairs may or may not depend on the means. If all the "natural" variables of the "anticipated state of affairs at any time" are completely independent of the "means" variables, although the internal state variables at this time may not be, such means are called *nonobjective*. They may determine the future values of the internal state variables of the actor. On the other extreme, if only the "natural" variables or actions at every future state are affected by the means variables at any time, and none of the future internal state-variables, such means are

called (purely) *objective*. Of course, the vast majority of means fall between these two extremes, and are simply called means. The terms objective and nonobjective, when applied both to means and ends, shall henceforth refer to some degree of proximity to the extremes, not the poles themselves.

If certain variables of the "natural" environment depend solely on their own previous values and the action of a particular actor, such variables are called *individually controlled means* relative to that actor (e.g., the position of a specified car is a means controlled by its driver, because the action of the driver and the present position of the car determines its position one time unit later). Clearly, there are controlled means which can be described for any actor (e.g., position of arms) and controlled means which refer only to a particular actor or type of actor (e.g., position of a 500 lb. dumbbell). Furthermore, there are means-variables which depend on the actions of more than a single actor (e.g., the position of a two-man saw in the middle of a heavy cut). In such cases, the proper distinction between means and conditions becomes more complicated, although the pair of actors who "control" such a means could be regarded as a single actor. Means which depend on more than one action shall be called "shared" means. The conditions of nature are specified by those variables of the "natural" environment which depend on no action, but only on the previous values of such variables.

Relations between means and ends

I. Logical Types of Action

An action shall be called of a *purely logical* type if:

(*a*) The future state of affairs to which the end-variables refer, whether it occurs or not, is objective. (i.e., no future internal state variables of the actor enter into its descrpition) and

(*b*) The associated means are objective.

To the extent to which internal state-variables are absent from the "anticipated state of affairs" and to which the internal state-variables of every future state of affairs are independent of the means, to that extent the action is objective (or empirical). The objective actions may be classified into two types:

A. A logical type of act is called adequately logical if, for a particular condition, (part of the situation) the means (assumed to be objective) are such that the ends (i.e., "anticipated state of affairs") are "close" to the actual future state of affairs.

B. If this is not the case, it is called inadequately logical, or illogical.

The degree of logical adequacy of an act is the probability that under the given conditions and given the means associated with the act, the actual future state will be "close" to the ends.

II. Nonlogical Types of Action:

There are three types of action in which the means and the "anticipated state of affairs" are not *both* objective to a large extent. They are:

1. The case where the means are nonobjective but the future state to which the ends refer is objective. It is possible that the means are such that the objective future state is "close" to the subjective ends with high probability, because even nonobjective (not purely) means codetermine the future values of the "natural" variables to a small extent. In this case, the act is called *adequately methodologically* nonlogical[5] if the corresponding probability is low, the same name applied with "inadequately" replacing "adequately."

2. If the means are nonobjective and the future state of affairs to which the

[5] In some work as yet unpublished Professor Omar Khayyam Moore of Yale University has raised an extremely interesting point in relation to some forms of magic. Professor Moore has suggested in a discussion of this work with one of the present authors that certain types of magic have the effect of randomizing search-type activities (e.g., the direction in which to look for fish). In the absence of scientifically tenable knowledge in such situations social action which randomizes efforts may well be a far better adjustment to the environment than any other. As such it may have very real survival value. In this sense without the actors concerned understanding what is involved the magical practices may well "work." This would be an ideal example of what is here termed "adequately methodologically nonlogical" action.

ends refer, are nonobjective, it is also possible to have adequately and inadequately, *ultimately* nonlogical action, depending on whether the means are such as to make the future values of the state variables "close" to the ends with high or low probability.

3. If the means are objective and the future state to which the ends refer is nonobjective (e.g., by means of certain drugs going into a trance wherein the actor believes he is Napoleon), such a type of action is called hallucinatory nonlogical. As in cases 1 and 2, it is possible to define a degree of adequacy. Heuristically, it appears that the degree of adequacy for cases 1 and 3 would be very low in most instances. For instance, although the means of prayer for rain might frequently be followed by the occurrence of rain, the effect on the probability of this happening is probably zero (case 1). On the other hand, prayer for happiness being followed by a state of happiness for the actor may not be so improbable (case 2). But again, the means of drugs to produce a specified internal state (say believing to communicate with ancestors) is not impossible, but at the present state of knowledge not a frequent occurrence (excepting, perhaps, drug addiction, etc.). Clearly, if the dependence of any future system states on the "means" variables were completely deterministic, adequacy for cases 1 and 3 would be meaningless, because a correspondence between the nonobjective means and objective ends or vice versa could not be stated with empirical meaning. Even in this case, it is possible for (erroneous) statements about such a relation to be coded in terms of certain of the actor's internal state variables.

The number of means refers to the number of combinations of values of the "means" variables which permit the occurrence of the ends. For every objective means, there is at least one future state of affairs objective or nonobjective to which the (subjective) ends may refer. The number of means for a given future state of affairs, however, may be 0, 1, 2, . . . , n. It is not necessary that for every objective means there corresponds at least one end (subjective), because, even though a future state of affairs to which it might refer, exists, it may not be possible to encode this state as a statement to be expressed in terms of the internal state variables, e.g., the actor may not be "aware" of the end toward which he is oriented[6].

Relation between logical and rational action

Although the terms "logical" and "rational," as applied to action schemes by Pareto, Parsons, Levy, etc., have been used synonymously, it may be useful to create a distinction. It may be noted that the same terms are also used quite differently in various works, such as those of Marschak (14), Simon (23), etc. The logical aspects of an act as they are defined here refers to the relationship between the means and the ends which are associated with the act, but not to the manner in which the "output" (syn: "act," "response") is selected by the actor. For any given situation, as specified by all the system variables, it is possible to explore the logical consequences of an act if sufficient information is available. It is also possible to attribute this capability to the actor, except that he may not be in complete possession of all the required information. This involves the assumption that the actor is endowed with the logical features of a computing mechanism, i.e., he has a "reasoning" facility. If such an actor-unit "selects" the "optimum" action according to some criterion for "optimum" by means of such a "reasoning" faculty, his action will be called rational. (This term is chosen to conform with the usage in game theory.)

According to this distinction, an act may be logical but not rational, if the particular means used do not conform to certain rules, although they are suited to the end. For instance, an act of buying may be called rational in a given social system only if the means are such as to buy at the smallest cost. Nevertheless, a buyer may buy without using means which conform to this standard, and his action, though logical, would be

[6] Note the relation of these considerations to such concepts as latent, manifest, UIR, and IUR functions and structures. See References 12, pp. 83–88, and 16, pp. 50–51.

called irrational relative to these rules. This suggests a qualifier to the term "rational", say R-rational, where R represents a set of rules which define rationality. As another example, consider the act of creating an artistic image or a scientific conjecture. If the end of such endeavors are assumed to be the expression of a certain understanding or truth, the action is often logical (i.e., the conjecture is, in fact, true) although not rational relative to the standards of, say, scientific method. (E.g., it was not deduced systematically, or did not result in any way from deliberate, concentrated, systematic thought.) It is probably true, however, that for any logical action, there is some equivalent action differing only in the means, which is R-rational for any R. The set of rules, R, may be regarded as a special case of norms; that is, they are stored by P as particular values of the norm-variable, S_5. It may be observed that the various criteria of decision theory, such as maximin, Laplace, Hurwitz, Bayes; etc, are particular cases of R. It is also noteworthy that consistency has always been, even if implicitly only, an essential characteristic of any definition of rationality. By consistency in action will be meant the absence of contradictions in the consequences of the chosen or contemplated means.

There is another ingredient which seems to be present in many sets of rules for rationality, namely: precedent. If the means associated with an action involve the search for a precedent and the computation of a decision based upon this precedent (if it has been "successful" do it again), then this action is often called rational. Clearly, it may turn out not to be logical. Actions which are R-rational according to this R can be said to conform with tradition. It is clear that the set of rules, R_c which emphasize consistency as a chief prerequisite for rationality might be inconsistent with a set of rules, R_t which emphasize conformity to tradition as a prerequisite for rationality. That is, there may be an action, logical or otherwise, which can be R_c-rational but not R_t-rational, or vice versa. It is thus suggested that rationality be regarded as a relative concept, and that a universal definition of rational action has not yet been found.

Relation to formal model of an action scheme

Subjective and objective ends. Most of the concepts which have been discussed in this part can be concisely translated into the symbolic language introduced by the authors in Parts I and II. There is little to be gained in so doing, because most of the notions, even in their verbal form, are stated with sufficient logical precision for the present purpose. There are a few points, however, which need further elucidation. First, the notion of the identity of the subjective and the objective ends, which is basic to the concept of logical action, refers to a relationship between the content of a statement (the subjective end) which is stored in coded form, and the actual state of the universe at a certain (future) time (the objective end.) The latter can be described as the set of values of the variables which describe the state of nature and the outputs of other actors at a certain time. The occurrence of such an event, however, has a subjective significance to the actor (interpretation or perception) which may be stored in coded form. The coded forms of the perception of an actual event and of an anticipated event are comparable, and it is in these terms that the notion of identity is to be interpreted.

Closeness. Secondly, the notion of "closeness" of the objective and subjective ends requires further elaboration. It is intended to mean the same thing as the concept of neighborhood in topology. In this connection, it is necessary to specify the nature of the basic space more precisely. This space consists of all possible statements in coded form which the actor can store. A point in this space is thus a sequence of symbols selected from a specified alphabet. Intuitively, it is desirable to call two points "close" if the content of the two statements which represent these points are "close." If there is a 1–1 correspondence between statements and events, the topological structure of the event-space can be induced into the statement-space. More generally, the correspondence of events to statements may

be regarded as a mapping of the event-space into the statement-space, so that more than one statement corresponds to a given event, and more than one event to a given statement. Requiring this mapping to be bicontinuous can, under certain conditions, induce a topological structure on the statement-space in terms of that of the event space which is assumed to be given.

An illustrative application

To make the connection between the concepts introduced here and the ones previously developed clearer, and also to show how they can be interpreted in a concrete, practical situation, it is desirable to formulate the following example.

Let P denote a consumer, and suppose that everything about P which is relevant to the problem can be described by the values of six variables at several instants of time:

(a) Let x^h be the action P can take at time h, and suppose that this variable has three values:

$x^h = a \equiv$ the decision to buy from A.
$x^h = b \equiv$ the decision to buy from B.
$x^h = c \equiv$ the decision not to buy, but to obtain information about prices.

(b) Let s_1^h denote the biological state of P, and suppose that it can either be α (healthy) or β (sick).

(c) Let s_2^h denote the state of information about the prices of A and B which P has achieved up to time h. Suppose that s_2^h consist of coded statements of the form: "The probability that the cost of acquiring more information does not exceed the difference in the prices is p."

(d) Let s_3^h denote the value to P, and suppose that it is either good (G) or bad (B).

(e) Let s_4^h be the end-variable; any of its values represent a statement about an "anticipated event" in coded form. Suppose that at time h, $s_4^h \equiv$ "Purchase of a good car at low cost."

(f) Let s_5^h represent the norms variable; any of its values denote a set of rules or constraint according to which behavior is regulated. For example, in this case, s_5^h denotes coded statements of the form "the purchase must be a legal transaction," "If $|x_A^h - x_B^h| > y^h$, x^h should be a_3," "If $x_A^h > x_B^h$, and $|x_A^h - x_B^h| \ll y^h$, x^h should be a^2."

$\bar{P}$ denotes the environment of P, and is taken to consist of the "natural" environment E, described by the variable y^h, and of two sellers A and B, who are described by their prices x_A^h and x_B^h and whose products are identical. Suppose that E refers to a means of acquiring information about x_A^h and x_B^h, and y^h is the cost associated with this at time h.

A "means" in this case is a sequence $\{y^h, y^{h+1}, \cdots, \}$. It is controlled by a sequence of decisions $\{x^h, x^{h+1}, \cdots, \}$. Clearly, $x^h = 0$ unless $x^h = a_3$. The sequence of acts or decisions consists of a run of a_3's, and may or may not terminate in a_1 or a_2. If the sequence does not terminate, or terminates with an a_3, so that no a_1 or a_2 occurs, the action is illogical, because the end was to buy a car, and this was not done. If the sequence terminates in a_1, and $x_A^h \gg x_B^h$, and both cars are equally good, the action would be called illogical even if P were ignorant of the large disparity in price.

If x^h is selected in accordance with the rules described by the values of s_4^h, (the set of these rules being denoted by R), then the action is called R-rational. If R is a proper set of rules, then any R-rational action is also logical, but not conversely. Thus, the decision to buy may be logical, but may have been arrived at by purely irrational means, e.g., tossing a coin. The norms of rationality, R, may vary from group to group. If k sets of such rules are known to prevail in a given group, and the distribution of $R_1, \cdots, R_k$ over the population is known, it may be possible to obtain a more general and possibly a more realistic approach to a number of problems in which the concept of rational behavior is essential.

SUMMARIZING DISCUSSION

In Part I it was maintained that models for the general theory of action have been left in a residual state with regard to the definition of the concept of social action.

The residual state of these models is made explicit by the arguments of the functions

1 and 2 (p. 269), which may include factors not accounted for by heredity or by the nonspecies environment. If the form of these functions can be specified, certain characteristics of social action (goal-orientedness, adaptiveness) can be realized in the present scheme. For example, if s_1^h is suitably interpreted as specifying the "anticipated state of affairs," it is possible to find behavior functions which will realize any such future state in certain present situations. Such behavior functions are called goal-oriented. It is also possible to define the extent to which any behavior function is goal-oriented. The property of a behavior function to "realize" any "anticipated state of affairs" in certain situations corresponds roughly to the goal attainment aspects of any social system as introduced by Parsons (20).

Secondly, the behavior function of P_1 can be said to have an *adaptive* property to the extent to which the situation is taken into account in action, with regard to both the biological survival of P_1 and the achievement of the goal. Thus, an adaptive and goal-oriented behavior function will realize any "anticipated state of affairs" in any situation. It may be adaptive and only slightly goal-oriented, or goal-oriented and only slightly adaptive. The properties of a behavior function being both goal-oriented and adaptive correspond roughly to the *directive correlation* between the action and the situation with respect to a focal condition (goal) and to a "coenetic" variable (the variables on which the action depends), as introduced by Sommerhoff (24).

Thirdly, inasmuch as s_1^h may specify more than one "goal," requiring an ordered sequence of acts, or the simultaneous action of several parts of P_1, it is fruitful to define an *integrative* or coordinative property of a behavior function. That is, by regarding P_1 as "organized into components" (9) with the action of P_1 as a composite of the actions of the components, it is possible to realize a larger variety of goals. The extent to which the components do not act at random, but in synchronism with all the other components, is the degree to which a behavior function for a group of components is integrated.

Finally, there are certain aspects of s_1^h which are interpreted as the emotional tensions and the learned and acquired norms according to which behavior may or may not be regulated. To the extent to which a behavior function takes these emotional tension and normative variables into account to maintain certain internal states of "satisfaction," to that extent does it conform to the fourth aspect which Parsons ascribes to a social action system, (latent tension-management and pattern maintenance).

In Part II we sought to establish a substantive basis for the definition of the concept. The effort here turned on the method by which cognition was introduced into a newborn member of the social system. For these purposes it was assumed that the neonate does not start out with any substantively defined concepts, i.e., without Kantian categories of the mind or any similar construct. It was assumed that the child has a capacity for recording events, i.e., a memory. It was also assumed that initially the child will make gestures, cry, etc., which are signals occurring at random from a parent's point of view. It was also assumed that there will be a parent, or someone *in loco parentis*, who will interact with the child and will attempt to achieve what the parent views as the satisfaction of the child. This parent will interpret certain actions on the child's part as indices of dissatisfaction. The parent will not, however, have perfect knowledge of the child's state and hence will lack perfect knowledge of what, from the parent's point of view, must be done to satisfy the child. The parent will, however, have some knowledge of the status of babies in general. Thus it is extremely unlikely that the parent, either for lack of motivation or lack of knowledge, will never make a response which will "satisfy" the child. It was also assumed that the child's general state is such that it can record the resolution or lack of resolution of a given state, e.g., the child is in some sense aware that having been hungry and having been given milk by the parent it is no longer hungry, or that having been hungry and not having been given milk it is still hungry.

The problem posed by the situation was

one of mutual learning. The parent needs to learn what given outputs of the child mean or rather to interpret them in such a way that the subsequent action of the parent in its effects on the child may be interpreted by the parent as resolving a given state of the child previously interpreted as unsatisfactory to the child. On the other hand, the child must learn to communicate with the parent in regard to the child's state and the suitability of the parent's action in relation to the child's state. Furthermore the child, at least, must learn from scratch as it were. That is to say, the child starts with no substantively defined concepts in its memory. Two types of learning, according to our model, develop. The first might be called search learning, or learning to learn. The second might be called substantive learning; i.e., learning the detailed definition of things and situations which are employed by the other individuals (in this case the parent) with whom the infant is in contact.

The first form of learning is a result of the factors mentioned above; especially of the interest of the parent, the empty memory of the infant, its biological state, and the ignorance of the parent. For the process of learning to go forward the infant must first learn to use efficiently at least one technique which the parent may interpret as rejection, or as indicating a state of dissatisfaction. It is this technique and others of similar effect which are here considered examples of "command of the generalized negative." If the parent always knew exactly what every state of his infant was and acted to satisfy it, the infant would learn nothing. If the parent never acted so as to satisfy the infant it would also learn nothing. It is from the interaction of a state that "knows" nothing in a substantive sense (the infant) and one that knows something, but not everything, and is anxious to please, that the initial learning takes place.

The example par excellence of the initial form of the "command of the generalized negative" is crying. It is our contention following in a sense from our model that one of the first things a baby learns to do efficiently is to cry in order to get things done. There is good reason to speculate that crying is extremely likely to get singled out in this fashion in any society regardless of variations in social patterns as among societies. Crying is, of course, not the only action that the newborn infant can make which can be interpreted by the parent as rejection, and so forth. Kicking motions, moving the head in such a way as to seek the breast or avoid it, waving of arms, etc., are others. Crying is unusually well suited for the purposes, however, for both biological and communications reasons. In the first place it can probably be shown in purely physiological terms that the continuation of many states of the infant, e.g., hunger, cold, etc., will sooner or later result in crying. It may even be possible for physiologists to show that crying is a great deal more likely to take place sooner or later given a continuation of these states than are other actions which the parent is likely to identify as rejection behavior. But even were this greater probability not the case, if crying were no more probable or even less probable than other possible actions, it would be extremely likely to be singled out for reinforcement as a search device if only because of its greater probability of efficiency as a means of attracting parental attention. The cry can usually be heard whenever one is close enough or in proper line of sight to detect motions of various sorts, but it can also be heard in a great many situations in which sight or other forms of detection of action are not likely to be effective. For these, and perhaps further investigation could produce additional bases, crying has been used as the example of the type of learning which first takes place. In this model the first thing infants learn to do is to attract parental responses more efficiently by crying. This and what gets built upon it become the initial examples of social action in the life history of any infant in any society, unless of course it becomes possible to indicate that some similar learning takes place while the infant is in the intrauterine stage.

There are several elements at work in the process of learning the "command of the generalized negative." The lack of storage of substantive concepts in the memory of the newborn infant makes any signal of the nature of its early states necessarily devoid of

detailed substantive content as far as the parent is concerned. The most an infant at this stage can do in many cases of its various states is to attract the attention of a parent who is anxious to please in such a way that the parent will interpret the infant as being displeased with some aspect of things as they are. The parent on the other hand starts out with some conceptions as to what the infant's states are likely to be, e.g., hunger, wet, cold, etc., and some conceptions as to how to satisfy them. Since the infant's signal is initially devoid of any detailed substantive content, the parent must choose amongst various possibilities. It is significant that because the infant has not as yet learned anything, the parent cannot use what might in other circumstances be far more efficient techniques. For example, if the parent has reason to believe that what he interprets as the infant's source of dissatisfaction is one of eight equiprobable possibilities, he cannot divide these into groups of four and ask the infant if it lies in group A or group B and so forth. Because of the infant's lack of stored elements, the parent must try the possibilities more or less one by one in some order chosen by the parent. The parent may be right the first time in many instances, but he is also certain to be wrong in many cases and more often wrong than would be the case if the infant's initial capacities made possible such scanning action as that mentioned above. This conjunction of events guarantees the infant the possibility of recording both successful and unsuccessful "tries," ensures that the "inefficient" stratagems to which the parent must resort can initially convey more information to the infant's memory than more efficient ones might, and also gives the parent opportunity to learn more about the states of the infant. Insofar as the line of model building followed here is fruitful, it is perhaps on the whole fortunate that parents are more ignorant of the states of their infants than they care to admit in some societies.

While the search technique is being learned and subsequent to some degree of mastery over it, the way is open to detailed substantive learning. The infant having gained parental attention can learn to associate breast or bottles as such or the terms for them even, with hunger states and so forth. In the body of the paper above it has been shown analytically that a stage is reached in such a process as this at which a given set of events is no longer dependent on the nature of the initial stage of the infant. At this point there are involved elements in the state of the infant that are not explicable in terms of the heredity of the infant and its nonhuman environment, but which can be specified in detail and their relation to the original setting factors shown at least in terms of their process of development. It is this that is used as the basis of the substantive definition of social action suggested here. In essence any action would be called social if it can be shown or presumed to contain a cognitive element introduced to the actors as a function of some combination of these two types of learning.

One may speculate a bit further, moreover. For one thing the actions of infants symbolized by such statements as "they always learn to say no before they learn to say yes" seem less mysterious. It is not a fifty-fifty probability that an infant may learn "command of the generalized affirmative" before he learns "command of the generalized negative"[7]. In a sense there is no such thing

[7] Quotation marks have been left around the phrase "command of the generalized negative" because it is not intended to convey that the infant understands the concept of negation as such or that it plans its use of it. This is in essence impossible without some general understanding of the affirmative concept as such. It is rather that the infant learns to use with greater and greater efficiency responses which the parent can and does interpret as rejection or general signs of displeasure. We have used one of these actions, crying, as an especially appropriate example of the whole lot (including such actions as turning away, pushing objects away, etc.) for reasons stated above. In essence the use of the term "command" is introduced only to indicate in verbal terms a use of a technique. The "efficient" use of the "command . . ." is intended to indicate use of the technique in such a way as to decrease consistently the probability of inappropriate responses on the part of the parent and hence increase the proportion of subsequent signals of satisfaction by the infant. One may speculate, however, that it is from just such origins that later the actual concepts of negation and affirmation arise.

as a "command of the generalized affirmative." To select by means of the affirmative one must know in substantive detail what it is that is wanted.

This knowledge is absent in newborn infants and is not immediately acquired. Even at much later time periods, use of the affirmative may be frustrated by lack of common definitions and concepts about the substantive details of things wanted on the part of two or more actors in the situation. This implies that all infants everywhere in a sense learn first to signal "no," and only learn to signal "yes" efficiently or its equivalent in any specific case after a very considerable amount of other learning has taken place. Furthermore, whether or not an infant learns "command of the generalized negative" is not a function of the permissiveness or nonpermissiveness of the culture in terms of which his parents live. That is to say, in very loose terms the infant does not learn to use "no" efficiently because parents are always saying "no" to the infant. This may affect the rate of the infant's learning "command of the generalized negative" efficiently, but in the case of normal infants and normal parents or even in cases of some considerable abnormality the infant will learn "command of the generalized negative" anyway.

Finally, there are two other implications of the above. In the first place, whenever an individual completely outruns the substantive concepts stored in his memory via the processes mentioned above, he must in essence resort to some further use of the search learning to get the increase of information that will make possible additions to his store of substantive concepts. In the second place, although concentration on cognition in this paper was originally merely a stratagem for singling out one avenue along which a substantive definition of social action within the limits set by heredity and nonspecies environment might be sought, it emerges as perhaps a great deal more interesting than that. To the extent that the argument used here is tenable, the strategic nature of cognition for general social analysis would be heavily emphasized. Both evaluations and emotional reactions emerge as in some sense special cases of cognition though the reverse is not the case. Furthermore this emphasis on cognition does not carry with it either a bias in the direction of a special emphasis on consciousness or on the logicality or attempted logicality of action. Because of recent advances in mathematics and other disciplines in handling communication and information problems, such a development may be quite fruitful. The field of cognition is peculiarly amenable to treatment in terms of these techniques. To the extent that cognition in this sense is strategic for the general analysis of social action, whether that action be logical or not or even conscious, the field of social action may be more broadly thrown open to the application of such techniques.

Finally in Part III an attempt was made to make operationally precise the meanings of "logical" and "rational" as applied to actions. It appears that it is useful to distinguish between the two terms, reserving for the former the area dealing with the analysis of the relation between the objective and the nonobjective characteristics of ends and means and for the latter the relation of actions to rules of conduct, norms, etc. specified in some way. It is found that actions can be observed, in fact are common in each of the emergent categories.

Mathematical notation was introduced mainly with the view of providing a symbolic framework within which the creative mathematician may be stimulated to fit future mathematical theories of action.

REFERENCES

1. Ach, N. *Analyse des Willens.* Berlin, Wien: Urban and Schwarzenberg, 1935.
2. Bühler, C. *Der Menschliche Lebenslauf als Psychologisches Problem.* Leipzig: S. Hirzel, 1933.
3. Bühler, K. *Die Krise der Psychologie.* Jena: Gustav Fischer, 1927.
4. Coleman, T. S. *A survey of mathematical models of small group behavior.* Bur. Appl. Soc. Res. T. R. No. 10. Columbia U., N. Y. 1954.
5. Flood, M. Game learning theory and some decision-making experiments. In R. M. Thrall *et al.* (Eds.) *Decision processes.* New York: Wiley, 1954.
6. Kleene, S. C. *RM-704.* Monica, Calif.: Rand Corp. Sta., 1952.

7. Kochen, M. An information theoretical model of organizations. *PGIT-4 Proc. IRE* Sept. 1954.
8. Kochen, M. Networks of multiple sequence transducers. Presented at Assoc. for Comput. Mach., Philadelphia, September, 1955.
9. Kochen, M. Organized systems with discrete information transfer. Publication forthcoming.
10. Lazarsfeld, P., Berelson, B., & McPhee, W. *Voting*. Chicago: Univ. Chicago Press, 1954.
11. Lazarsfeld, P. (Ed.) *Mathematical thinking in the social sciences*. Glencoe, Ill.: Free Press, 1954.
12. Levy, M. J. Jr. *The structure of society*. Princeton: Princeton Univ. Press, 1952.
13. Lewin, K. *A dynamic theory of personality. Selected papers from Adams and Zener*. New York: McGraw-Hill, 1935.
14. Marschak, J. Towards an economic theory of organization and information. Ch. XIV. In R. M. Thrall *et al.* (Eds.), *Decision processes*. New York: Wiley, 1954.
15. McCulloch, W. S., & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* 1943, 5, 115–133.
16. Merton, R. K. *Social theory and social structure*. Glencoe, Ill.: Free Press, 1949.
17. Murray, F. J. Mechanism and robots. *J. Comput. Mach. Assn.* 1955, 2, 61–82.
18. Oettinger, A. G. Programming a digital computer to learn. *Phil. Mag.*, 1952, 43, 1243–1263.
19. Pareto, V. *The mind and society*. New York: Harcourt, Brace, 1935.
20. Parsons, T. *Towards a general theory of action*. Cambridge: Harvard U. Press, 1951.
21. Radbruch, G. *Der Handlungsbegriff in seiner Bedeutung für das Strafredits-system*. Berlin: J. Guttentag, 1903.
22. Shannon, C. (Ed.) *Theory of automata*, Ann. Math. Princeton: Princeton Univ. Press. In press.
23. Simon, H. A. A behavioral model of rational choice. *Quart. J. Econ.*, 1955, 49, 99–118.
24. Sommerhoff, G. *Analytical biology*. London: Oxford U. Press, 1950.
25. Turing, A. On computable numbers with applications to the Entscheidungs problem. *Proc. London Math. Soc., Series 2*, 1936, 42, 230–265.
26. von Neumann, J. Probabilistic logics. Cal. Tech. Lectures, 1952.
27. von Neumann, J., & Morgenstern, O. *Theory of games and economic behavior*. Princeton: Princeton Univ. Press, 1947.
28. Wald, A. *Decision functions*. New York: Wiley, 1948.
29. Weber, M. *The methodology of the social sciences*. (Tr. by E. A. Shils & H. A. Finch), Glencoe, Ill.: Free Press, 1949.)
30. Zelditch, M. Jr. A note on the analysis of equilibrium systems. Appendix B Parsons, T. and Bales, R. B. *Family*. Glencoe, Ill.: Free Press, 1955.

(Manuscript received March 15, 1956)

General propositions do not decide concrete cases. The decision will depend on a judgment or intuition more subtle than any articulate major premise.

—Oliver Wendell Holmes, Jr.

OBSERVATIONS ON SOCIAL PROCESSES IN PSYCHIATRIC RESEARCH

Definitions of Knowledge, Method, and Science in Psychiatry[1]

by Stewart E. Perry

National Institute of Mental Health, Bethesda, Maryland

Some characteristics of the social structure of psychiatric research seem to influence the psychiatrist's definition of knowledge and how it is obtained in his field. Certain experiences tend to set up questions in his mind about the scientific respectability of the common psychiatric clinical operations. These challenge the psychiatrist's self-picture as researcher and turn his attention from problems of science and method in psychiatry to problems of his role acceptability within the social situation in which he does his research. The result may be definitions of psychiatric knowledge and research methods which are more appropriate to the demands of the social situation than to the requirements of a psychiatric science.

Intellectual concepts of the nature of knowledge are defined within a social context that follows and is followed by the shape and perhaps even the content of the concepts. That is, man's definition of knowledge will vary from one era to another and from one society to another in association with variations in the social context of the definition. This is a general thesis in the sociological study of knowledge and of the roles and institutions that perform knowledge-defining functions within a society. In this paper I will try to describe the social processes that are associated with intellectual concepts that American psychiatrists, as researchers, use in the definition of knowledge of human behavior. The thesis I will present is that psychiatrists *under certain social conditions* tend to devalue their own clinical operations, in terms of the desirable qualities of science and knowledge, and to overvalue the operations of the quantitatively oriented social scientist.[2] A secondary thesis is that this process of evaluation takes place without as much attentive relevance to the intellectual or philosophical significance of the scientific operations involved as is given to the requirements of the social situation.

The formulations of this paper are based upon observations, direct and indirect, from four sources. (*a*) The primary source of data has been the author's participation in the psychiatric program of a clinical research hospital. In the course of a continuing study of the social organization of the clinical research program the author has made observations at general staff meetings held for administrative, clinical, or research purposes; and has conducted individual interviews with staff members to discuss the problems involved in the role-definition of

[1] I want to make acknowledgement of the help of my colleagues in clinical research in adult and child psychiatry at NIMH, who have discussed with me the role of the psychiatrist as researcher. I also want to thank Phillipe V. Cardon, Erving Goffman, Helen S. Perry, and Charlotte G. Schwartz for helpful comments, suggestions, and discussions.

[2] The author's value judgment on this should be made clear: I feel that this tendency is unfortunate; that it is destructive of good psychiatris research effort of the nonquantitative varietiec without bringing about a commensurate, balancing increase in the effectiveness of quantitative psychiatric research.

the psychiatrist as researcher. Intensive studies of the natural history of a number of individual and group projects were also made. (*b*) The author's experience in other clinical settings, where research was only an occasional or peripheral function of the staff psychiatrist, affords some contrast to the research setting. (*c*) Further experiences in a psychiatric training institute which also develops and supports research projects in the interdisciplinary study of human behavior have provided contacts with psychiatrists doing research outside of a hospital setting. (*d*) Finally, the anecdotal and other casual reports of psychiatrists about their experiences in other institutions around the country have afforded some check on the particularity of the author's own experiences.

The question of the generality of these observations and the formulations derived from them should be dealt with here before proceeding further. First of all, some of the discussion is relevant primarily to the segment of the psychiatric profession which uses psychotherapeutic theory and techniques—especially the psychoanalytic orientation. But this will be clear in context, and the relationship of the intellectual orientation of that segment of the profession to the orientation of other groups within psychiatry will be indicated. It would seem to be appropriate to think of current American psychiatry in terms of psychotherapeutic psychiatry, especially as it has been influenced by psychoanalytic theory, since it can be fairly safely assumed that such is the current dominant theme of American psychiatry. This would mean then that although I shall be talking in terms of psychotherapeutic psychiatry, presumably what I have to say will be applicable in terms of the psychiatric profession rather generally.

Second, although I confine my attention in this paper to psychiatry and its problems of defining method, knowledge, and science, the thesis I advance in terms of psychiatry might be advanced equally in terms of other branches of social science.[3] Especially psychology and sociology have been dealing with the question of quantification for perhaps a century. Indeed sociology with its inception through Comte began immediately to deal with the question of scientific stature in relation to quantification. Psychiatry, however, has only recently become involved in this sort of questioning. Perhaps the reasons for this have to do with psychiatry's isolation from the social sciences as a medical discipline and with its only recently increased theme of psychologization in the past fifty years or so.

A more general intellectual trend which may currently keep the question of quantification in the foreground of psychiatry and the other social sciences is the tendency toward interdisciplinary investigation and theory construction which focuses attention upon the techniques and methodologies of other disciplines.[4] But a far more important factor in the social science confusion over how to know things about human relations is probably the broader social context in which the social scientist operates: the American value system which gives a high place to engineering and technology and their associated techniques of investigation. It is interesting to contrast American social science in general with its European counterpart in this respect. In Europe—as the most cursory examination of the literature would suggest—library, qualitative, and deductive studies in all the social sciences command a greater respect than they do in the United States, where inductive, quantitative studies are a more developed work area.[5] It is interesting to note, in regard to

[3] A parallel discussion, but in terms of his own field, is provided by Lewis Atherton. See Reference 1.

[4] Much recent discussion has taken the form of examining the difficulties and uses of multidisciplined collaboration. See References 2, 3, 5, 11, 15.

[5] Thus an American psychiatrist of some repute was recently led to say of another equally well-known psychiatrist: "I have heard from those who have no liking for him that he does his research in the library." Incidentally, the gap between Europe and the United States in respect to acceptable scientific investigation may be narrowing with continuing relationships between scientists of the two continents. It would seem likely, however, that in view of the material advantages of United States research, the gap may tend to be narrowed more by European social science becoming closer to the United States, than by the United States position moving across to

the United States bent toward quantification, that Comte, a mathematician as well as a sociologist, decried in his own day the emphasis upon mathematical description as the identifying quality of scientific work; he thought such an attitude was born of materialism (6). These broader social considerations will not be dealt with further, although they represent one level of analysis of the social definition of knowledge. The succeeding discussion will be confined to an analysis based upon the more direct, more easily distinguished, and more particular kinds of social experiences that the psychiatrist undergoes in some performances of his knowledge-defining functions as a researcher.

It might actually be said that the psychiatrist in his ordinary relations with his patient performs a general epistemological and ontological function. That is, his operations with the patient are in terms of defining what is real and what he and the patient need to know. He is concerned with mapping out with the patient the latter's areas of difficulty in living, and with clarifying with him what is and is not, in the patient's life—all this to the end that with a clearer idea of his own participation in his problems in living, the patient will be able to solve them. This aspect of the psychiatrist's role, however, is not immediately socially relevant in terms of the definition of psychiatric knowledge, although in the performance of this function the psychiatrist may gain knowledge which he will then formulate for society at large in papers, books, and addresses. It is the goal of formulating information for a wider audience than the particular patient that makes the psychiatrist a knowledge-definer in the sense of contributing to a field of knowledge. Thus, to study the psychiatrist's social experiences and how they are related to his definition of psychiatric knowledge, one must look at his activity where it is more directly involved in producing, receiving, or transmitting knowledge in this wider sense. This means that one must be concerned not with his service or purely therapy functions but his research, learning, or teaching activities.

the European. However, narrowing from both directions will undoubtedly take place.

As a matter of convenience for discussing the social processes involved in the psychiatric definition of knowledge, I will proceed by describing more or less in developmental terms the way a psychiatrist moves into and accepts a knowledge-defining role as a researcher. This sort of career-line description has advantages in providing some contrast in the transition from a service orientation to a research orientation, when the psychiatrist accepts a job in which he is to do research.[6] Learning and teaching as other knowledge-related activities will be described tangentially to this transition.

In tracing the development and definition of a research role by the psychiatrist, three groups of social processes will be described; each group corresponds roughly to a stage in the genesis of the psychiatrist's definition of psychiatric knowledge insofar as this definition is conditioned or accompanied by relevant social processes. The first group of processes is that inherent in the psychiatrist's training and experience before he assumes a research role. The second group of relevant processes has to do with the expectations engendered by the actual research situation. The third group of processes are those reactions or responses to research which the psychiatrist may see in his associates and which he may emulate and use in his own definition of knowledge, method, and research.

PSYCHIATRIC TRAINING AND EXPERIENCE

There are at least three features of the psychiatrist's background and training in psychiatry which seem to have importance in his later definition of psychiatric knowledge as a researcher: (*a*) He tends to have little or no academic exposure to research methodology or the philosophy of science. (*b*) The opinion of other segments of the medical profession about the status of psychiatry as a scientific discipline is not high. (*c*) There is the common practice in medicine of doing research with patients in terms of ordinary clinical procedures that would be

[6] This approach has the further advantage of staying quite close to the primary source of data for this paper—that is, the author's study of the initiation and continuing organization of a psychiatric research program.

used for treatment purposes anyway—a practice leading to a certain lack of intellectual distinction between clinical procedures and research. Each of these three features of the psychiatrist's background may be seen simply in terms of their being experiences which tend to set up accompanying scientific ideologies and professional values.

Training

It will have to be granted that the psychiatrist who goes through the ordinary training course in America rarely, if ever, receives any training specifically in research philosophy or methods. Even if he goes on to take a good deal of advanced training, such as passing through the course of a psychoanalytic institute, he is not required to pay much attention to the research and scientific status of his tools, techniques, and philosophy.[7] In short, he learns under intensive training how to practice psychiatry and psychoanalysis, but he is not led to understand self-consciously the methodological and scientific underpinnings of his profession, as medicine in general and as psychiatry in particular. Although typically he will not have been familiarized with research operations in his psychiatric training, the psychiatrist may, if he has long planned to be a researcher, seek special training for that purpose before taking a research job. But the fact that the psychiatrist's apprenticeship in medical school, residencies, and psychoanalytic training is so prolonged militates against his further prolonging his education by special training for research (9).

The psychiatrist, then, who takes on the role of researcher—such as specifically accepting a job as a researcher—is apt to come to it relatively naive as to how he will perform his research functions. He will understand, to be sure, that his training background is the reason he has been hired, and so he expects or presumes that the techniques and approaches he has been accustomed to use in previous work will be the ones he will use in a research job. His expectation may be either supported or challenged by his work group when he assumes his new position—a point which will be discussed in a moment.

A final comment on the part played by the learning experience of the psychiatrist should be made. Psychiatric training is, at least in its earlier stages, medical training; therefore it imbibes of the long medical tradition of empiricism. Medical empiricism has tended to isolate the doctor from interest in systematic theory. Such isolation is, of course, a handicap to pursuing a research career (12).

Interprofession standing

Another characteristically professional value-idea-experience complex which the psychiatrist brings to his research job is his peculiar role as a specialist within the medical profession. This experience is, so far as most psychiatrists are concerned, remarkably similar: that is, most psychiatrists feel that their specialty is regarded by other medical students or practitioners as a most unreliable and diffuse discipline. In the view of other medical men, psychiatry may be something that is rather strange and even anxiety-provoking, but most of all it is simply a step-relative—not a real relative—of the medical science and profession. The

[7] For example, according to their recent (1954 and 1955) training catalogs, a number of the American Psychoanalytic Association training institutes have no class time devoted to the subject of scientific method; others may have 6–24 class hours listed as devoted to methodology, but course syllabi would indicate that the listing is somewhat inappropriate. A couple of exceptions to this picture are institutes devoting 20–48 required class hours to the examination of research methods and the place of psychoanalytic techniques as compared with other sorts of research techniques. Other medical psychiatric training centers follow the same pattern as the psychoanalytic institutes. Incidentally, it should be noted that the main goal of such training centers is equipping persons for treatment functions; equipping students for research functions may be an academic luxury not easily afforded. It is interesting to note in conjunction with the above facts that the standards for training of psychoanalysts, as promulgated by the American Psychoanalytic Association, does set up as a required part of the institute course curriculums training in scientific method:

"Psychoanalysis is presented as a method and as a synthesis of empirical observations.

"Special consideration is given to primary postulates, to basic requirements for utilizing the method of free associations and to logical means for interpretation. Applications of criteria to clinical material provide continuity with the course on dreams" (4).

psychiatrist is thus almost an outsider so far as the other medical men are concerned (13, 14). To be sure he is held in higher esteem than the osteopath—after all, a psychiatrist does have an M.D. degree[8]—but the respectability of his professional subject matter is looked at a bit askance. This question of respectability is not so imperative if the psychiatrist defines his interest as neurological and physiological aspects of psychiatry. Defined thus, psychiatry is closer to traditional medicine in practice and theory. If, on the other hand, the psychiatrist is primarily concerned with the psychological and interpersonal aspects of human living, his subject matter is far removed from the medical, and it demands a great deal of understanding for his medical confreres to accept his specialty as a "real" part of medicine. For one thing, behavior as defined psychologically and interpersonally is less amenable to many of the common techniques of medical research. That is to say, the psychiatrist cannot operate very easily in a laboratory; he tends instead to operate in a clinical, person-to-person relationship with the patients whom he is studying. The heart specialist, the ear specialist, and so forth, have two avenues open to them for their medical investigation: the laboratory method and the person-to-person or clinical method. The psychiatrist must usually be content with but one approach to his data. He must be a clinician par excellence. In fact, his professional standing within his specialty will depend upon his powers of sensitive perception in his relationship with his patients. The difficulties of conducting psychiatric research (of the mentalistic or psychological type) according to laboratory techniques analogous to other medical research techniques, therefore, tend to prevent the psychiatrist from using a standard means to demonstrate the validity of his frame of reference as scientific in comparison with other medical fields.

[8] The emphasis that medical psychotherapists place upon the medical degree in their relations with nonmedical psychotherapists may be related to a need to maintain a certain status under this adverse pressure from other medical colleagues.

Merging treatment and research

The research-psychiatrist-to-be brings also to his research role a certain confusion between service and research. The clinical method—that is, everything that is involved in diagnosis and treatment—is service, and it is engaged in for the purpose of service to the patient, his family, and the community, in the ordinary practice of medicine. Most doctors are in service positions, there being more demand for service than for research. It is not unusual then for the medical man interested in research to carry on his research on a bootleg basis, sandwiched in between his service functions. Research carried on this way must be done ever so economically and there is therefore the added impetus to do the research with the identical procedures which will be used anyway in the treatment function. Research thus may not be differentiated from the purely service or clinical function of the doctor—except in terms of the additional value placed on the treatment procedures by the doctor for himself. Psychiatrists have shared with other medical men this sort of global experience of treatment-cum-research. An indication of the close relationship of ordinary clinical treatment routines and research in medicine (especially newer branches of medicine) is provided in the literature of single case studies and studies of very small numbers of cases.

THE RESEARCH SITUATION

With this sort of background the psychiatrist approaches the situation in which he consciously sets out to do research in a group setting.[9] The central expectation which, like any other researcher, the psychiatrist experiences is, of course, the demand that he produce research results. This is a demand which may be more implicit in basic research, more explicit in applied research; in either case, *results* mean basically

[9] For all practical purposes—in terms of the sociology of psychiatric knowledge—psychiatric research not carried out in a group setting probably is ultimately no different, although the private researcher may be isolated to some degree and for some time from influences more powerful in a group setting.

some sort of scholarly or professional publication or some systematic verbal presentation of one's work to one's peers—a presentation that is susceptible to publication if there be no extraneous reasons for it to remain unpublished.

This demand for results means that the worker is expected at some point not too long postponed to stand up and say, "This is the final report on what I have been doing; this is the result, the reward to science (and/or to me and to my employer), for the time I have spent and the work I have done." The researcher in any field is required to put himself out on a limb, and so too the psychiatrist. This demand is so much a part of research that it is to be presumed that a researcher at least tolerates it easily or even accepts it with alacrity, so long as he feels he is capable of fulfilling the expectation. It is only when the researcher feels an impotence, real or fantasied, in the face of the expectation or demand, that he will resist it. At the point of growing resistance to the demand for results there is an increase in the worker's counter-demands for the resources and facilities that may be useful for meeting the demand for results. For example, there may be increasing requests for supporting personnel and equipment or for additional salary. If these requests are granted, there is a corresponding increase in the original demand that sparked the request for more support. There is a cyclical process: The more equipment, personnel, space, or what-not that the researcher has placed at his command, the more responsibility he is likely to feel in terms of using these facilities to get results. There is thus a rather inescapable complex of interlocking expectations—and physical buttresses to those expectations—which reinforces the whole research system, so that the worker must develop ways of maintaining his own confidence in his ability to meet these expectations, if he is to remain a researcher.[10]

The maintenance of one's self-picture as a researcher is important for the purposes of this paper in that in at least psychiatric research one of the maintenance processes involves the definition of scientific knowledge in the research field. Before discussing this point, however, it is important to probe the psychiatric research setting for some of the elements that will tend to threaten the maintenance of an acceptable self-picture and thus call up a social adjustment which has intellectual consequences. For this purpose let us consider, as the typical setting, the multidisciplined setting in which most psychiatric research currently takes place. By this I do not mean that most psychiatric research is done necessarily in collaboration with other disciplines, but merely that the research is done in close physical or intellectual proximity to research going on in other disciplines or with other disciplines. This fact is due primarily to the growth of group research rather than individual research. Psychiatrists tend to do research in institutional settings or in collaboration with others who are in institutional settings, and therefore are more apt to be in some kind of contact with workers in other disciplines. For example, if the institution is a general hospital, there will be other medical specialties. Universities, research organizations, and mental hospitals are also loci for specialists in other social or medical disciplines. The private practitioner as a lone-wolf researcher is almost unknown in these days, and even if a psychiatrist operates by himself physically, he would find it difficult to avoid the intellectual influence of interdisciplinary work done by his psychiatric colleagues. The interpenetration of

[10] Tremendous pressures are built up in the individual researcher which can take a physical and psychological toll. When one is given everything in equipment, assisting personnel, and salary, in return for which mere miracles are expected, results must come out in terms of miracles or in terms of symptoms, as Nellis Foster has observed in banking and business situations, as well as in medical research. This "Nellis Foster syndrome" is reported by Alan Gregg in an address, "The Natural History of Changing Your Job," delivered at the National Institutes of Health, U. S. Public Health Service, September 30, 1954, as one of the 1954 series of National Institutes of Health Guest Lectures, proceedings to be published. See also Reference 7.

the social sciences is thus an almost inescapable intellectual influence.

In the multidisciplined psychiatric setting the psychiatrist comes in contact with social scientists[11] who apparently are studying the same sorts of items of human behavior that the psychiatrist studies. However, for studying the same thing the social scientists are able to choose from among a greater assortment of tools than the psychiatrist has at his disposal. First, there are those techniques and tools of which the psychiatrist may not have any intimate knowledge and which the social scientist may flaunt before him. The techniques of experimental science, adapted or maladapted by social sciences from the arsenal of the physical sciences, are a case in point. These techniques, generally quantitative and thus sometimes requiring advanced mathematical knowledge, are reminiscent of the laboratory medical sciences, and they may therefore face the psychiatrist with the same stern, hawk-nosed visage that he saw within the medical pecking order. Second, some of the most advanced training a psychiatrist can claim a social scientist may also have. For example, there is the important experience of personal analysis which the social scientist is quite likely to have undergone if he is interested in research that brings him into much contact with a psychiatric setting. The social scientist may even have undergone systematic psychoanalytic training as a therapist—although, of course, such training would generally have taken place outside of classical psychoanalytic institutes. In other words, the psychiatrist may experience difficulty maintaining the necessary self-picture as a researcher because (*a*) he does not have certain research tools of high social value that are available to his associates and/or (*b*) he has no individual tools, no tools peculiar to his own profession at his disposal.[12] This contrasts sharply with the ordinary work situation of the psychiatrist in a clinic or hospital, and particularly in private practice, where his own functions are considered by himself and by nonmedical associates as much the most important, where he receives a greater differential in terms of financial return and in terms of authoritative standing, and where his operations are closer to being socially unique.

Examples in formal and informal social behavior

At this juncture I should like to give a few illustrations of the point that psychiatrists find it necessary to make some socio-intellectual arrangement or integration of the contrast between clinical operations in research and operations which are closer to laboratory, experimental, and mathematical models. One almost across-the-board experience is the evaluation in research and other medical institutions of the neurological, brain-metabolical, and so on, studies as being somehow more "basic" than studies which cannot use the laboratory methods. For example, in the hospital here under consideration, two branches of research are organized, each organizationally on a par with the other. One is named Clinical Investigations; the other is named Basic Research. The branch called Basic Research is, as one might expect, the branch which is most concerned with laboratory techniques and with neurophysiology, neurochemistry, and the like. The Clinical Investigations branch, on the other hand, deals mostly with actual patients and is the branch that uses less the laboratory or experimental method than the naturalistic method. It is interesting to note a further implication in the differences between the

[11] Although psychiatry can be considered one of the social sciences, I use the term *social scientist* here to refer to psychologists, anthropologists, sociologists, and the like.

[12] The purely medical training which distinguishes the psychiatrist from nonmedical practitioners in his field is, he will sometimes assert, relatively useless to him in most instances in the practice of psychotherapy or in the study of the interpersonal processes of behavior with which psychotherapy is concerned. On the other hand, for ceremonial and other purposes emphasis may be placed upon the relationship between the psychotherapeutics and medicine—an emphasis which serves to strengthen the respectability of psychotherapy. A presidential address to the American Psychoanalytic Association begins: "Psychoanalysis is biologically based and body bound. Historically, its most intimate relationship has been as a medical discipline . . ." (8).

two names. *Investigations* is a much less rigorous term in science than is *research* (10). Further, within the Clinical Investigations program there is one unit which is less concerned with mentalistic or interpersonal investigation than it is with physiological functioning. This unit, interestingly enough, is termed the Laboratory of Clinical Sciences, whereas the unit which is most concerned with psychotherapy with adults is termed the Laboratory of Adult Psychiatric Investigations. The difference in the names here is even more striking.

These organizational titles, of course, not only reflect the underlying intellectual philosophy of research and science, they may also support it. Although the public evidence of this intellectual orientation may not be as striking in, for instance, university settings, still those who have been connected with them can testify to similar relative scientific standings of subject matter and methods of systematic inquiry.

At this point let me take some more informal illustrations: A group of researchers (including sociologists, psychologists, and psychiatrists) discussing their work over coffee was asked by one of its members for advice on placing an article already completed. A certain psychiatric journal was suggested by a sociologist, but the reply was, "I mean a scientific journal—you know, one that publishes papers with statistics." Although the particular journal in question does in fact regularly publish "papers with statistics," it undoubtedly publishes less of that sort of material than others; for that reason it is not a "scientific journal!" A research psychiatrist studying psychotherapy was talking about his concern over criticism of his work: "Those guys [psychologists] may have something when they talk about 'reliability,' but all I know is that I have to do my kind of work." But was his kind of work *science*, he wondered. Another psychiatrist—with a considerable reputation—addressing a heterogeneous group of researchers said, "Of course we can't prove scientifically that psychotherapy works—that is, we can't proceed like Koch in bacteriology—but I intend to spend the rest of my professional life in a commitment to study psychotherapy, in the faith that there is something to it." In spite of his undoubted capacity and reputation, he apparently felt it necessary to note apologetically that his work was more a matter of faith than science.

The author himself had a prolonged and agreeable association with two psychiatrists on a project in which however he found himself in an anomalous position: The relations of the three were troubled primarily by the expectation that the author would somehow make the original and significant naturalistic observations of the two psychiatrists into something quantitative and scientific. The author had the continuous role of pointing out the greater significance of the observations by the psychiatrists, in comparison with some minor manipulations of the data by statistical means, which the author had made at their suggestion. This difference of opinion was never adequately resolved. (Perhaps it will be with this paper.)

A final example involves the use of humor as a response to the felt demand that psychiatric research follow the course of laboratory, experimental research. A waggish psychiatrist put his finger neatly on the sensitivities of psychiatry in a research hospital by greeting his colleagues in the corridors with the following remark: "Discovered anything recently?"

Comparison of psychiatric and other research

The psychiatrist, then, is more or less forced to make adjustments to the technological arsenal of the social scientists (or laboratory scientists) in the multidisciplined research setting because he (or others) cannot escape making comparisons of the gross scientific standing of his own work and that of his associates. The comparison that the psychiatrist makes may be invidious in terms of (*a*) techniques; (*b*) results; and (*c*) criteria to evaluate such techniques and results. That is to say, the psychiatrist may begin to assume that his techniques are not as sophisticated as the instruments of the psychologist, for example. Both his techniques and their results, he knows, are less precise and less communicable, operationally speaking. Moreover, the very criteria

with which he is accustomed to evaluate his own hunches and work are suspect. For the psychiatrist is apt to judge his clinical hypotheses and results by the following yardsticks: (*a*) Did the patient respond in the direction of improvement? In other words, did the psychiatrist's operations succeed empirically? (*b*) Do the various facts brought out in the course of treatment from all sources—from the patient and from those with whom the patient has had contact—fit together into a relatively coherent pattern? In other words, is there a tendency toward mutual confirmation of all data? The first criterion is, as the psychiatrist himself will admit, susceptible to self-fulfilling prophecy; moreover, "cure" or "improvement" is a proof the definition of which few psychiatrists agree upon. Thus in terms of standards of reliability it does not stand very high as a criterion of the scientific status of a piece of work. The second criterion—mutual confirmation or coherence of data—is a relatively respectable yardstick, so far as the history of science is concerned. However, its use in psychiatry requires that one assume a large body of untested and vulnerable theoretical formulations. That is, various items of data can be considered coherent and compatible only within a context of a theory that can manage them. No current psychiatric theory is noted for its widespread uncritical acceptance, and thus data may fit into a theory and still not receive much credence. The psychiatrist, viewing his own standards thus may come to the conclusion that in comparison with the chi squares, correlation coefficients, and probability scores of quantified social science research his own work looks rather weak.[13]

What also makes the comparisons between himself and social science inescapable is the fact that in the sphere of the psychological-interpersonal the psychiatrist's subject matter is very similar to the interests of the psychologist or social anthropologist, for example. He is threatened by eclipsing forces in, as it were, his own heaven. The threat often appears to him to come through the superior power of the social science techniques. Such a threat is presumably important, because there is fundamental competition within a research field, just as much as in the real estate business or commercial baseball. That is, the fruits of successful competition are professional standing or advancement, which apart from its economic advantages is a basic source of one's self-picture in an achievement-oriented society. The competition within a research program does not simply take the form of seeing who will do his work better than whom. It can take many sorts of forms. One of these is especially marked by irrationality: staking out claims in the lands beyond the horizon of current knowledge and maintaining such property against potential claim-jumpers—as if somehow the wealth of unsolved scientific problems could be exhausted. Thus certain areas of investigation may be jealously guarded. There are more obvious forms of competition in terms of amassing facilities that investigators use in their work: equipment of any sort, but especially expensive equipment, and supporting personnel, either junior assistants or senior consultants.

In the psychiatric research situation the competitive game tends to suggest to the psychiatrist that he adopt an aggrandizing attitude toward research problems, facilities, and techniques. Instead of sharing his interests he will indicate they are closed to outsiders. Instead of being challenged and thus presumably satisfied as a scientist with the examination of his clinical, interpersonal operations with the patients whom he studies, he looks for other fields to conquer. Instead of the usual clinical facilities, he seeks out intricate equipment to observe the patient: The first prerequisite to psychiatric research, but only the first, is the high fidelity tape recorder and the one-way observation screen. And he looks for collaborators or technicians who can provide him with the key to Science. In a sense then, the psychiatrist may come to see cer-

[13] It is ironic that the psychiatrist can become as dazzled as he sometimes does with the paraphernalia of precision. With rare exceptions, one cannot say that the paraphernalia have been used to brilliant advantage in the social sciences.

tain forms of science as being the substance of science. He begins to equate the tools of quantification with the finished product of a competent piece of investigation. He seems to consider the science of behavior studies not as an unfolding development within and without his civilization but instead as whatever social scientists currently do. He may even begin to see the science of behavior studies as what the social scientists of his acquaintance do. In short, research becomes inextricably confused with someone else's technology.

THE RANGE OF RESPONSES

I have discussed up to this point the kinds of experiences which set up pressures toward the undervaluation of the common psychiatric clinical operations as scientifically respectable. Now I should like to describe some of the socio-intellectual adjustments which the psychiatrist makes within the research situation to a perception of the relative position of his own operations versus the more quantitatively expressed operations of other researchers. There are a number of ways by which the psychiatrist can make an adjustment to this perception which poses a potential challenge to his self-picture. As he enters a research institute for example, he can choose from among the response processes he sees in his fellow workers who have established a modus vivendi. Among these response processes are some which maintain the necessary self-picture by adjustments having relevance to the definition of psychiatric knowledge, the fruits of the research operation. I would like to describe briefly two such processes: The first of these may be called *conversion*; the second, *orthodoxy*. These terms are chosen to emphasize the social nature of the knowledge-defining response, as opposed to the intellectual content of such a response.

In a response characterized by the process of conversion the psychiatrist sees the true light of science as quantification. He looks back upon psychiatry's clinical, naturalistic inquiries as crude fumblings. As a convert, he vows to follow the call of pure Science ever after. To this end, he may seek to learn the quantitative and other skills or information he has not picked up in medical training.[14] Or he may attach himself to possessors of such skills hoping he can contribute some ideas that can be studied in sterilized laboratory detail by his purer associates. Or he may limit his research to problems that are amenable to the statistical description and the laboratory instrumentation that he can execute by himself. In any case he considers the clinical method too crude to use, and he may treat a few patients "only in order to keep my hand in," for purposes he considers to be unrelated to his research interests. The conversion process results in the definition of psychiatric research in terms not very distinguishable from other behavioral research, because it is defined in terms of their means. Conversion ends in the rejection of the traditional psychiatric clinical procedures with individual patients in favor of nomothetic methods; and it rejects the validity of information which churns around in the psychiatric brain rather than in the mechanical calculating machine.

A quite different process that may be present in the psychiatrist's definition of his research role, and derivatively present in his definition of psychiatric knowledge is the process I have called orthodoxy. It is antithetic to conversion. Its fundamental intellectual premise is that psychiatry is a discipline *sui generis*, a discipline which should not be contaminated by the introduction of heretical concepts or technologies derived from the social sciences. Psychiatry in terms of this process comprises phenomena far removed from the ordinary subject matter of social science; it is in fine the study of mental aberrants, people who need to be helped. The study can be undertaken only by someone with the sixth, or clinical, sense; by the same token, the clinical understanding of what is necessary for the well-

[14] Doctors are generally required to take academic courses in mathematics, either in premedical training or in medical school. The application, however, of this academic knowledge is ignored in his training, and it may be as useless to him for research purposes as having studied the bony structure of the foot is to the practicing psychoanalytic psychiatrist.

being of the patient is a *sine qua non*. Orthodoxy thus commits the psychiatrist to the study of mentalistic phenomena for the most part; it also commits him to therapeutic goals in the course of his study.[15] In short, he is primarily interested in what goes on between him and the patient in the course of treatment, and he may thus be concerned with the so-called intuitional processes.

Orthodoxy is regarded as magical thinking, by the psychiatrist who has been converted to quantification. In turn, orthodoxy suggests to the apostate convert that there are more things in heaven and earth than are dreamed of in his statistical orientation. The antithesis between conversion and orthodoxy may be expressed by a member of a research program in the following train of thought: "Quantification does not respect the individuality and uniqueness of the whole person; therapy requires such respect; on therapeutic grounds, therefore, the sorts of procedures necessary for quantified research results are out of order." Orthodoxy then can provide a rationale for the psychiatrist to refuse to allow his patients to be used for validating a psychological test, or for any other presumably quantitative procedure.

The two processes may thus be found in conflict with each other; and within a psychiatric research program, individual staff members may come to symbolize for each other these typical processes. The two processes are found, to be sure, in any individual staff member who assumes the occupational role of psychiatric researcher. They may, therefore, result in a synthesis which might be called *reform*. This sort of synthesis is signalled when the psychiatrist tries to reconcile both methodological schools, and does so simply on the rationale that "both approaches have something to offer." The reform process is expressed when the psychiatrist delivers the opinion that clinical investigations comprise a phase in the broad development of science, or comprise a necessary, though rough section of the road that leads to investigations by more exacting techniques and with more precise results. The reformist response is, in short, vaguely broad-minded and tolerant. In this tolerance the psychiatrist is free to use the techniques and concepts of his past training, but he must also be deeply interested in interdisciplinary work. The psychiatrist does not usually have time to give for interdisciplinary work himself, but he must couch his reports to colleagues in terms that show a sensitivity to multidisciplined feelings. He may be somewhat humble and suggest that he does not aspire to be a precise researcher, but he feels you will agree he has a "darn good" therapy project.

These then are three of the possible processes in the psychiatrist's response to the new demands that a research career places upon the clinician in him.[16] Before discussing a fourth possible response process, I want to examine some similarities of the responses already described. As patterns of response, these processes in the psychiatrist's definition of his field are marked by a singular absence of any examination of the meaning and structure of the methodologies he uses or those he envies. These processes are, in fact socially-conditioned definitions of knowledge, method, and research in psychiatry, but they are lacking in intellectual content, so to speak. That is, the internal logic of such definitions is not developed.

As social definitions, these points of view or responses to the research situation can develop into divergent schools of so-called thought within the same research program. Such schools of thought are marked more by social isolation from each other than by intellectual boundaries. As time goes on, however, the social isolation *may* be accompanied by the development of truly idiosyncratic techniques that are not tem-

[15] On the other hand, the psychiatrist's prior commitment to therapeutic and mentalistic interests may influence him in his choice of the *orthodox* point of view on the science of psychiatry.

[16] There are other sorts of responses which are irrelevant to the present discussion on the sociology of psychiatric knowledge. For example, the psychiatrist may become an administrator rather than do studies of his own, or he may return to private practice. In either case, he has to a great extent withdrawn from a role that requires he explicitly define psychiatric knowledge.

pered by the interest and contact of other points of view.

There is, of course, another sort of response—though in my experience a more rare response to the research situation. A psychiatrist may figuratively sit himself down to puzzle out exactly where his particular clinical operations come into the picture of scientific methodology. Such a response process is as social as any of the other three. That is, it is stimulated by the necessity to find a socio-intellectual position within the structure of a specific behavioral research institution and within the structure of the research community in general. What distinguishes this response from the other three are the following characteristics: (*a*) A definition of psychiatric knowledge is self-consciously evolved; it is not merely an uncritical adoption of the epistemological and methodological assumptions of other behavioral researchers—as is the case with the conversion process. (*b*) An attempt is made to see the psychiatrist's clinical operations as fundamentally the same as any other scientific means of inquiry; there is no attempt to separate psychiatry as a *sui generis* discipline or set of procedures. (*c*) And finally, instead of merely accepting, as given, the close relationship of psychiatric techniques of investigation to techniques used by other disciplines, the psychiatrist explicitly attempts to discern the interrelations of the varying approaches and to integrate them in terms of their scientific meaningfulness.

Although the psychiatrist may thus attempt to develop an intellectually-defensible rationale for his operations, unless this rationale can receive acceptance within the social situation of the research work group,[17] he may find it an inadequate adjustment to the need to paint the picture of himself as a scientific researcher. His methodological position, like others within his particular setting, will probably proceed toward its own logical extreme in inverse ratio to the capacity of adherents of the other positions to understand its fundamental intellectual structure. In the process of this development toward extremes, the social acceptance of the definitions may take the form of the development of methodological schools of thought, as previously noted, and the schools themselves may become the significant feature of the research group atmosphere rather than the content of the methodologies. For the content, or internal, intellectual structure of the scientific position of frame of reference, must in such cases be free of examination, else the whole social adjustive structure collapse. That is, the definition of acceptable method and of scientific knowledge in psychiatric research will depend upon what the social structure of the research setting requires, rather than upon what the logical nature of subject matter for research might dictate.

Some characteristics of the social structure of psychiatric research seem to influence the psychiatrist's definition of knowledge and how it is obtained. The socially-conditioned definitions appear to overvalue the methods of quantitation—positively or negatively—and undervalue the methods of clinical investigation. Because the need for a rationalized methodological position in a research setting is highly social, the psychiatrist's attention is focused on what will meet the social need, sometimes to the point of ignoring what will meet the scientific or methodological need. The responses that he makes to the social situation tend to become institutionalized and there is likely to be a decreasing attention to questions of scientific method for itself. The conclusion one is apt to reach is that the clinician requires a developed and communicated scientific rationale for the problems he chooses to study and for the methods those problems require. The institutionalization of such a rationale in the early stages of psychiatric training and by the development of clinical research programs will provide the psychiatrist with an internalizable frame of

[17] To a great extent the likelihood that this response pattern is used will depend upon such potential acceptance. Particularly it will be encouraged by favorable attitudes toward the naturalistic research of clinical observation held by those members of the staff who are prestige figures. For example, the continued encouragement of the top research director can, to a certain extent, offset unfavorable attitudes held by what might be called the researcher's peer group.

reference that is not only intellectually sound but socially defensible.

REFERENCES

1. Atherton, L. The research center in entrepreneurial history: a personal appraisal. *Explorations in entrepreneurial history*, 1954, 7, 105–110.
2. Blackwell, G. W. Multidisciplinary team research. *Soc. Forces*, 1954–1955, 33, 367–374.
3. Bronfenbrenner, U., & Devereux, E. C. Interdisciplinary planning for team research on constructive community behavior. *Hum. Relat.*, 1952, 5, 187–203.
4. *Bulletin American Psychoanalytic Association*. Standards for the training of physicians in psychoanalysis, 1950, 6, 1–5.
5. Caudill, W., & Roberts, B. H. Pitfalls in the organization of interdisciplinary research. *Hum. Organization*, 1951, 10, 12–15.
6. Comte, A. *Politique positive*, Vol. I. In Ward, L. F. *Pure sociology*. New York: Macmillan, 1903. P. 47.
7. Gregg, A. When to change jobs—and why. *Harper's Magazine*, August 1955, 211, 71–76.
8. Kaufman, M. R. Psychoanalysis in medicine. *Bull. Amer. Psychoanal. Assoc.*, 1951, 7, 1–12.
9. Kubie, L. S. Some unsolved problems of the scientific career. *Amer. Scientist*, 1953, 41, 596–613; 1954, 42, 104–112.
10. Lee, A. M. Individual and organizational research in sociology. *Amer. sociol. Rev.*, 1951, 16, 701–707.
11. Redlich, F. C., & Brody, E. Emotional problems in interdisciplinary research. *Psychiatry*, 1955, 18, 233–239.
12. Shryock R. H. *The development of modern medicine*. New York: Knopf, 1947.
13. Smith, H. L. Psychiatry: a social institution in process. *Soc. Forces*, 1954–1955, 33, 310–316.
14. Smith, H. L. The relationship of psychiatry to the other medical professions. Paper read at Amer. Sociol. Soc., Washington, D. C., 1955.
15. Wohl, R. H. Some observations on the social organization of interdisciplinary social science research. *Soc. Forces*, 1954–1955, 33, 374–383.

(Manuscript received March 1, 1956.)

The speculative philosopher and the scientific specialist are liable to two opposite mistakes. The former tends to deliver frontal attacks on Reality as a whole, armed only with a few wide general principles, and to neglect to isolate and master in detail particular problems. The latter tends to forget that he has violently abstracted one part or one aspect of Reality from the rest, and to imagine that the success which this abstraction has given him within a limited field justifies him in taking the principles which hold therein as the whole truth about the whole world. The one cannot see the trees for the wood, and the other cannot see the wood for the trees. The result of both kinds of mistake is the same, viz., to produce philosophical theories which may be self-consistent but which must be described as "silly."

—C. D. BROAD

DISCUSSION AND REVIEWS

THREE SYMPOSIA

The interdisciplinary symposium on behavioral theory has become within the last decade a common occurrence in American academic life. The pattern was set largely by the well-known Conferences on Cybernetics sponsored by the Macy Foundation. Indeed, what is known today as "Behavioral Science" stems from the cooperative effort of workers who constitute the usual cast of characters of similar conferences: representatives from psychology, the social sciences, biology, communication engineering, and mathematics. The following are reviews of three such conferences. Two of them, "Problems in Human Communication and Control" and "Information Theory in Psychology," revolve largely about the same focal point of interest, namely information theory and its relation to behavior, particularly to problems of communication and control. The third, "Toward a Unified Theory of Human Behavior", is broader in scope, in that the approaches to behavior characteristic of psychiatry, anthropology, and the social sciences (in which the conceptualizations usually differ largely from those of quantified science) are equally emphasized. The levels of discourse of the published material, on which these reviews are based, are also different. The National Science Foundation Conference is a single conference of a few days' duration and is presented as a transcript of a recording with only enough editing to smooth out the inevitable roughness of very lively oral exchange of views. The material of the Illinois conference (also a single conference) is published as a collection of formal papers interspersed with running comment by the editor. "Toward a Unified Theory of Human Behavior," representing four conferences held over two years, is a rather thoroughly edited account in which the contributions have been organized around more or less formal oral presentations followed by discussion. It is reviewed with interspersed critical evaluations by a representative of another group which over a prolonged time has wrestled with related problems and who consequently sympathetically recognizes at once the difficulties and the potential contributions of such interactions. In reviewing the two conferences on information theory and allied topics, we have concentrated our own editorial comments in the following remarks.

The Promise and Pitfalls of Information Theory

INFORMATION theory, repeatedly hailed as a "major breakthrough" came at a time when mathematical sciences had just become newsworthy—on the heels of nuclear fission. But while the impact of the suddenly publicized atomic physics has been on the entire population through the conventional channels of popular fancy—the awesome possibilities for good and evil of vast amounts of suddenly harnessed energy—the appeal of information theory has been largely confined to academic circles. Yet the nature of this appeal has been not unlike the other. It is felt that a new conceptualization has crystallized which will lead to momentous theoretical reconstructions in

biology, psychology, and the social sciences. A vast proliferation of papers devoted to attempts to exploit the new conceptualization reflects this feeling. A 1953 bibliography of information theory[1] contains some 800 entries. Writings on the subject range from sophisticated analysis of radar systems and television circuits to crackpot speculations. Usually, there is a well-marked negative correlation between the scope and the soundness of the writings, so that it is not hard to understand the disappointment with "information theory" as a broad conceptual tool: the sound work is confined either to engineering or to rather trivial applications (rote learning, etc.); ambitious formulations remain vague.

The creators of information theory could, of course, point out that their discipline was conceived in the context of communication engineering, computing machinery, and automation, that they did not coin the phrase, "Life feeds on negative entropy;" that the parallel between entropy and information has been noted merely as a formal identity of mathematical expressions, and that those who wish to speculate about the effects of the Second Law of Thermodynamics on a "closed society" and similar matters should do so at their own risk, etc.

Yet the challenge of extending the concepts of information theory remains, and, as a matter of fact, is traceable to the writings of its founders.[2] It seems useful, therefore, to try to see what the problems of extension and generalization actually are, that is, to try to state the issues as "problèmes bien posés."

Given a set of configurations of any system with associated independent probabilities of occurrence p_i, the *uncertainty* of the set is defined in information theory as

$$H = -\sum_i p_i \log_2 p_i \qquad [1]$$

[1] F. L. Stumpers, *A Bibliography of Information Theory*. Cambridge, Mass.: Research Laboratory of Electronics, Massachusetts Institute of Technology, 1953.

[2] Perhaps a historical analogy is found in the proliferation of speculations concerning the broader meanings of the "uncertainty principle" in quantum theory. Actually many of these speculations are found in the writings of Bohr himself, one of the fathers of quantum physics.

Thus if one selects a message from a source of n messages, each selection is a "configuration" characterized by a certain probability. Then H is the uncertainty (per message) associated with the source. The receipt of the message transmitted without error "destroys" the uncertainty of the recipient, with regard to which message will be chosen. Therefore H measures also the amount of information per message.

Now in statistical mechanics, the thermodynamic quantity *entropy* also appears as the same expression, where now the set of configurations comprises all the arrangements of positions and momenta of the particles comprising a system that correspond to the thermodynamic *state* of the system. It turns out that states described in thermodynamics as *equilibrium states* have corresponding to them the greatest number of such configurations and so are "most probable." The trend toward equilibrium observed in all systems "left to themselves" is interpreted in statistical mechanics as the predominance of occurrence of the more probable states over the less probable.

A homely example illustrates the principle. In a well-shuffled deck of cards the number of red cards among the first 26 cards will be not too far off from 13. At any rate with continued shuffling, numbers around 13 for reds among the first will appear far more frequently than numbers around 26 or 0. This is so because many more arrangements of the deck are possible where the reds are evenly distributed than when they are all in one half. It is justifiable to say that the "equilibrium condition" (where a red card has as much chance of passing from the first half of the deck to the second as vice versa) is the condition of the greatest "entropy" of the deck.

The condition "if the system is left to itself" is crucial. Nothing is easier than to lower the "entropy" of a deck of cards by deliberate re-arrangements. Once thermodynamic entropy was recognized as a measure of "disorder" in a system it was natural to raise the question of whether the Second Law could be circumvented by the intervention of an "intelligence." The seed of the idea relating information to entropy

was sown. The idea first appears in the writings of none other than James Clerk Maxwell himself, one of the founders of statistical mechanics. He writes in *Theory of Heat* "... a being whose faculties are so sharpened that he can follow every molecule in his course ... would be able to do what is at present impossible to us....Let us suppose that a vessel is divided into two portions A and B by a division in which there is a small hole, and that a being who can see the individual molecules opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower ones to pass from B to A. He will, thus, without expenditure of work raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics."[3]

Thus was born Maxwell's Demon, possibly a descendant of the Homunculus to challenge by means of his phenomenal acuity and infinitesimal reaction time the inexorable Second Law.

Maxwell's Demon, or Maxie, as we will call him for short, now 85 years old, is still a subject of serious theoretical discussion. A paper of historical significance for information theory was devoted entirely to him by L. Szilard[4]. Szilard has shown that in lowering the entropy of a gas by his "decisions" Maxie causes an increase of entropy *elsewhere* which, as calculations show, more than compensates for the decrease and thus vindicates the dictum of the Second Law that the entropy of a *closed* system (which in this case must include Maxie and his apparatus) can only increase. More recently, L. Brillouin[5] emphasized Maxwell's assumption that Maxie *can see* the molecules. Thus he cannot be sitting in the total darkness of the enclosed gas (even if the gas is hot and radiates light, it doesn't help him to distinguish the molecules if the whole system is in equilibrium). To see the molecules Maxie must use a flashlight and thus, as, Brillouin shows, feeds "negative entropy" in the form of a light beam into the system. It is this negative entropy which is more than an ample source to account for the resulting decreased entropy of the gas alone.

Thus Maxie is shown to be cheating like any charlatan exhibiting a perpetual motion machine. These discussions are interesting, because they may lead to insight concerning the way enzymes work. It is the enzymes which are credited with the "local" contraventions of the Second Law which seem to be operating in living systems. On the other hand, it is known that eventually all enzymes become "denatured" and lose their function as "organizers" of living processes. ("For an enzyme to be in equilibrium is to be denatured."—Wiener) It has been remarked that besides counting the rise in entropy in the gadgets which Maxie must have for his entropy-lowering activity, one should take into account the rise in entropy within Maxie himself, perhaps the denaturation of his enzymes reflecting the "entropic costs" of his decisions. This, in turn, leads to speculations concerning similar processes in all "organizing" activity, the perfection of technology, accumulation of scientific knowledge, etc. Are we becoming "denatured" to the extent that we "order" the world to suit our needs? We mention this only to illustrate how easy it is to embark on metaphysical and even theological speculations once one becomes seriously concerned with the Second Law.

So much concerning the possible exploitation of the conjectured equivalence of information and physical entropy. Let us turn to the speculations revolving around the concept of information as a tool for quantifying the "amount of knowledge." Here, even more than in matters dealing with physical entropy the scope of attempted generalizations far outruns available tools of analysis. This is not surprising, since "knowledge" evokes strong intuitively felt similarity to information and since knowledge has been traditionally the subject matter of philosophic speculation. Ac-

[3] Quoted by L. Brillouin, *Journal of Applied Physics*, 22, p. 334 (1951).

[4] L. Szilard, Über die Entropieverminderung in einem Thermodynamischen System bei Eingriffen intelligenter Wesen," *Zeitschrift für Physik*, 53, 840-856, 1929.

[5] L. Brillouin, "Maxwell's Demon Cannot Operate: Information and Entropy. I," *Journal of Applied Physics*, 22, 334-337, 1951.

cordingly, one finds a wealth of discussion about fighting the inexorable Second Law by increasing our knowledge (knowledge = information = negative entropy), about the inevitable collapse of totalitarian societies (totalitarian society = closed society = closed system, in which entropy is bound to increase) and similar matters much removed from the theory originally designed to inquire into conditions which make for efficient transmission of signals over a channel.

We must disavow in passing any suggestion that we consider such speculation as entirely idle. After all, the Copernican revolution had a profound effect much beyond astronomy, and the law of conservation of matter has implications much more far reaching than making for correct bookkeeping in the chemist's laboratory. However, the profound effect of celestial mechanics and the conservation laws on our world view operated, we think it is safe to say, through the resulting success of scientific method, and the latter depends for its operation on precise definitions of situations one deals with. Therefore it seems to us that the search for applications of information theory to communication in a broader sense than it is understood by those who think in terms of wires and tubes (i.e., broadly to problems of human communication), while commendable and attractive, must be tempered by a sober attempt to keep the universe of discourse well defined.

A difficult transition is from the concept of information in the technical (communication engineering) sense to the semantic (theory of meaning) sense. This transition has been undertaken by Carnap, Bar-Hillel, and other workers. The difficulties involved become immediately apparent. Suppose we ask how much information is contained in an average English word. Has the question any meaning? This depends on to whom it is addressed. The telegraph engineer has a most definitive answer. He knows that the redundancy of English reduces the amount of information per letter from about 5 to 2 bits, and since the "average" word has about 5 letters, this makes 10 bits the information content of the average word. The telephone engineer's or the linguist's answer requires a more intricate analysis in terms of phonemes, distinctive features, etc., but his answer can, *in principle*, be equally direct.

Now what does 10 bits per word mean? It should mean that if I think of an English word and you try to guess it, you should be able to do it in about 10 guesses requiring yes or no answers. If you choose your guesses judiciously, you should be able to do it simply guessing letter by letter. But suppose letter guessing is disallowed. You are required to guess, as in the game of Twenty Questions, by narrowing down logical categories. As people who play Twenty Questions know, extremely bizarre notions can be guessed within that limit; so it is possible that an experienced player can guess a single word in ten tries. It is remarkable, however, that while it would be rather simple to program a logical computer so as to guess any word letter by letter, to build a machine that would do so by "logical categories" would involve building into the machine the entire encyclopedia of human knowledge! Moreover, any eight-year-old can be taught to guess any word letter by letter (even unfamiliar words, if the question, "Is that all?" is included). But if the word is not in his repertoire, no amount of advance instruction will enable an eight-year-old to guess it by "logical categories."

This illustrates a cardinal feature of information theory. It is fundamentally a theory of *selection*. Something is selected from a well-defined set. To examine the selective process, it is essential to be able to examine the set.

The modest but, we think, significant applications of information theory to certain psychological experiments owe their success to the fact that in each situation the set in question was strictly defined: a list of syllables to be memorized, associations to be formed, responses selected from, etc. There was, therefore, no difficulty in quantifying the associated "amounts of information" and relating such accounts to certain aspects of performance.

To many these results appear somewhat

trivial. We do not agree. If information theory is to break out of its original habitat of bandwidths and modulations, a proper beginning must be made, which usually means a modest beginning. Experimental psychology has made this beginning possible. However, one must admit that the gap between this sort of experimentation and questions concerning the "flow of information" through human communication channels is enormous. So far no theory exists, to our knowledge, which attributes any sort of unambiguous measure to this "flow." It is naive to take simply the flux in signals per second to multiply by bits per signals in the communication engineering sense and call the result "amount of communication" in the sense of transmission of knowledge (labeling everything one does not like "noise"). If there is such a thing as semantic information, it is based on an entirely different kind of "repertoire," which itself may be different for each recipient. To be defined as a quantity of information a signal must be selected from a set or matched with an element of a set. It is misleading in a crucial sense to view "information" as something that can be poured into an empty vessel, like a fluid or even like energy.

Yet some meaning lurks in the expression "to acquire information"! We feel, however vaguely, that as a race we have learned certain things as a result of which we can more effectively "order the universe" to our liking. We have learned to reverse the degradation of energy locally by putting heat to work. We like to think that in increasing our average life span, we are challenging (again only locally) the trend toward chaos (all living things seem to do so, but we think we do it better). At this point we could be accused of the same promiscuous speculation we have implicitly warned against. We can only plead that such speculation is extremely difficult to avoid.

Let us try for a while to walk the tight rope between rigor and speculative abandon. It was probably the now famous remark of Schroedinger, "Life feeds on negative entropy," that has been responsible for speculations which are always tantalizing but which range from profound to irresponsible. A most prosaic interpretation is simply in terms of thermodynamic bookkeeping. The First Law having long been established, it is axiomatic that life feeds on energy: it converts chemical energy of ingested food and of sunlight into work necessary to pursue its activities. The energy equation must balance. However, with the advent of the Second Law in its classical, i.e., purely thermodynamic form, another bookkeeping appears necessary. The entropy inequality must hold. A living organism must ingest at least enough *free* energy, which contains a negative entropy term, to account for any local decrease of entropy concomitant to its *organizational* activity. All this would be perfectly clear if the "increment of organization" inherent in a process were as clearly defined as an increment of energy. One might argue as many do, that organization is simply the negative of entropy. True, entropy is unambiguously defined in classical thermodynamics, but this discipline confines itself only to processes which take place under continued equilibrium (reversible processes). The classical definition, therefore, is inappropriate for living processes ("For an organism to be in equilibrium is to be dead."—Wiener). The advent of statistical mechanics with its re-definition of entropy as the probability of an aggregate of states does make possible the extension of the entropy concept to non-equilibrium states, indeed in information-theoretical terms. Hence the excitement. One feels it may be possible to calculate the "amount of organization" of configurations arising in the living process and thus establish the bookkeeping demanded by the Second Law to deal with those aspects of life, which, it is felt, are much more crucial than those covered by energetics bookkeeping (First Law).

One such attempt, to give an example, is due to R. C. Raymond[6]. Raymond proposes to define the entropy of a configuration by the sum of two terms, a positive one,

[6] R. C. Raymond, "Communication, Entropy, and Life," *American Scientist*, 38, 273–278, 1950.

equal to the entropy the aggregate would finally attain if left to itself (i.e., its classical entropy) and a negative term equal to the amount of information necessary to reconstruct the original configuration from equilibrium (from complete chaos). This definition is attractive and indeed fits perfectly into the situation previously discussed where Maxie works with 100% efficiency pumping information (= negative entropy) into a system which starts with positive entropy only.

We suspect, however, that there are grave difficulties associated with this definition when one tries to extend it to situations other than the simplest (ideal gas, etc.). The difficulty is the same conceptual one which governs the applications of information theory to semantic problems. What are the elements of a configuration? And what is the set from which messages describing the construction of a configuration from chaos are selected? That is not to say that it is impossible to define these elements, but only that it appears that the definition must be made anew in each instance. How then is one to agree on a definition once for all so as to make the entropy of a configuration unambiguous?

Let us take an example. How much "order" is contained in a particular arrangement of a deck of cards? According to one view, since each arrangement is equiprobable its probability being $(52!)^{-1}$, it follows that $\log_2 (52!)$ yes-no decisions are required to reconstruct a particular order from a thoroughly shuffled deck. Now although (52!) is prodigiously large ($\sim 10^{68}$) its logarithm is not (~ 200), and it is perfectly feasible to guess the arrangement in about 200 guesses, provided one is allowed to guess card by card. But suppose one is to guess by arrangements only? *We simply do not have the language* to classify the card arrangements so as to be able to make judicious dichotomies for yes-no answers. Our knowledge of card arrangements is confined to very few, which we call "ordered" (e.g., arrangements by suits and face values), and we tend to lump all the other arrangements as "meaningless" chaos. Similarly, a dictionary is useful only because its entries are arranged in a *particular* order meaningful to us. It is no consolation to be told that any other arrangement is equally improbable and therefore should be considered as having an equally low "entropy." One suspects that similar considerations apply to biology and make Raymond's definition of order less useful than it appears. Only *certain* arrangements maintain life. It is misleading therefore to measure the amount of order simply by the amount of talking one has to do to give instructions for its reconstruction or by how improbable the configuration in question is. All *particular* configurations are a priori equally improbable. As for the amount of information necessary for reconstruction of order, paradoxically the arrangements which we tend to consider ordered are usually reconstructible with *fewer* instructions. To reconstruct the ordered card arrangement, we need only say, "arrange in descending order of values the spades, clubs, diamonds, and hearts," whereas to reconstruct an arbitrary arrangement from another arbitrary one, we will generally have to name the position of each card separately! So in biological situations one might think certain arrangements have to be first *singled out* as the significant ones. Then one can hope to quantify the degree to which an arrangement approaches the "proper" one. But the discovery of "proper" arrangements is not simply a matter of counting; it is a matter of understanding the life process. Hence much more than understanding the theoretical tool of information theory is required for a proper application of the tool.

With regard to the other proposed extension of the entropy concept (as it appears in information theory), i.e., to the study of human communication (intertwined with semantic problems), the task is at least as difficult. To answer the question, "How much does a man know?" or "How much has he learned?" is possible in the sense of the present state of information theory only if one can specify a set from which selections must be made. Information is measured by the extent to which it makes uncertain selection certain. An attempt to define such repertoires of selection often leads to conclusions which seem, to say the

least, unsatisfactory. Suppose, for example, a person does not know whether ghosts exist. Then he reads a book which presents weighty evidence against the existence of ghosts, as a result of which he becomes convinced that ghosts do not exist. How much information has he acquired? One can "reason" as follows. Prior to reading the book he was equally uncertain about which answer is proper, "Yes" or "No." After reading the book, he is certain of the answer "No." Therefore he has acquired one bit of information. We submit, the conclusion is unsatisfactory. (We will refrain from raising the question about the case where the man is convinced by reading a book that ghosts do exist.) Or take another example. A child learns that $2 + 2 = 4$. How much has he learned? The formal argument could run thus. If the child is acquainted with the names of numbers 1, 2, ... n and was equally uncertain about which was the right answer, then he has acquired $\log_2 n$ bits of information. This conclusion seems to be equally ludicrous. These examples are given solely to illustrate what pitfalls await those who wish to extend the formal aspects of information theory to applications with more human interest. It is tempting to suggest that the American Revolution was triggered by one bit of information ("One if by land; two if by sea"). In this case the contention is not wholly unjustified: it refers to the simplicity of the arrangement for implementing one of two alternative plans of resistance, depending on the choice of the British commander to move by land or by sea against the insurgents. The simplicity of the arrangement was made possible by previous elaborate organization (of Minutemen, etc.) which it served to trigger. But such reductions of complicated situations to well-specified alternatives are rare.

Lest we seem to pause on a pessimistic note, let us reiterate our faith in the future of information theory. It is sometimes instructive to learn from historical analogy. A precursor of information theory, of which the latter is an elaborate extension, was probability theory, which started from humble beginnings as in intellectual schematization of what happens in games of chance, and which has grown to become one of the fundamental tools of exact science, whose logical foundations are a challenge to most serious philosophers. We note that attempts to extend probability theory to epistemological problems, for example, were also studded with pitfalls. No less a figure than Laplace, starting with a plausible theory of induction, came to rather ridiculous conclusions concerning the probability that the sun would rise on the following day. The axiomatization of probability theory is of recent origin and rests squarely on the definition of a "sample space," i.e., an explicit listing of the range of events over which the probability measure is to be defined. The notion of sample space seemed to have given a *coup de grâce* to the hope of extending exact probabilistic notions to measure a "degree of belief," a subjective concept. (What is the probability that the Democrats will win in 1956?) Yet there were those who refused to give up hope of successful axiomatization of "personal probability," so as to extend the strictly objectivistic view adhered to by the majority. As a result we have, for example, the highly interesting formulation of L. J. Savage[7] in which this axiomatization is achieved, not without the use of rather sophisticated conceptual and mathematical machinery.

We believe that extensions of information theory and of the entropy concept, which is now riding on the coat tails of information theory, are entirely in order. We are afraid, however, that these extensions are no easy matter and cannot be achieved by "semantic suggestion," that is, a tendency to confuse invention of nomenclature with discovery of principles. (*AR*)

Problems in Human Communication and Control. Paraphrased transcription of a conference sponsored by the National Science Foundation, June 15–17, 1954, at the Massachusetts Institute of Technology. J. C. R. Licklider, Editor. (Mimeographed)

[7] L. J. Savage, *Foundations of Statistics*, New York: John Wiley & Sons, 1954.

By way of channeling the discussion into the main current of thought on the subject of concern, J. C. R. Licklider opened the conference by summarizing the sessions on Information Theory in Psychology at the International Congress of Psychology in Montreal. In particular, the broad topics dealing with the application of information theory in psychology were restated: first, the study of a human being as a channel through which information flows; second, the problem of pattern in sequences of symbols as represented by language and other symbolic processes; third, the problem of measuring the "amount of organization" in any situation, an especially challenging problem when the usual objective quantitative characteristics (dimensionality, metric scales, etc.) are absent.

Thereupon the discussion centered mainly on the first of these topics, the channel capacity of the human operator.

Henry Quastler described the nature of the simplest experiments in which the subject appears as a channel. Stimuli are selected from a source characterized by a certain entropy (defined in the usual sense pertaining to discrete sources). In the simplest case, a one-to-one correspondence exists between the stimuli and the appropriate responses. The rate of presentation of stimuli, then, determines the rate of information input. As long as the subject responds correctly, his rate of information output equals that of the input. If the latter increases, so does the former until at some rate errors begin to appear.

The typical input-output curve is then monotone increasing with decreasing slope and horizontal asymptote. This asymptote, which depends essentially on the information rate (rather than the complexity of the task) can be interpreted as the "channel capacity." Thus three pianists of high caliber reached an output of about 22 bits/sec. Assuming 6.5 bits per English word, capacity for English text for typists appears to be about 24 bits/sec. However, for random type script, it is something like 15 bits/sec.

Discussion then centered about the question of why it is advantageous to express these rates in information measures. Quastler pointed out that the principal value of the information measure is that of an aid in generalizing. If it turns out that the capacity is characteristic not of a specific task but of a class of tasks (within which the tasks may vary in complexity, rate of presentations, etc.), then it would be possible to have "something like a catalogue of human capabilities in performing various tasks" from which the performance on a new particular task could be predicted from information-theoretical considerations.

Norbert Wiener cautioned about the difficulty of determining the input and the output associated with a human channel in view of the great importance of information storage and the utilization of stored information in the performances of humans. He also brought out an analogy between looking at a "span" of material (as in reading) and the action of a wave filter. Filters, it appears, can be designed more efficiently if a considerable amount of lag can be allowed.

Another set of experiments involving the individual as a channel were described by Irwin Pollack. Here the center of interest was not the rate of transmission but the amount of information transmitted per stimulus presentation as measured by the accuracy of reproduction. For example, the subject is required to reproduce the position of a dot in a square, a linear pattern of dots, randomly arranged letters or numerals, etc. Other experiments involved the ability to identify speakers from samples of speech (presumably by characteristics of voice, not content), etc.

Some general principles seemed to emerge. There was a maximum amount of information associated from a single "dimension" of display (see below the discussion of the meaning of "dimension" in this context). Multiplying the "dimensions" allowed the transmission of more information, sometimes proportionately. The increase was noted up to six or eight dimensions.

An argument then arose on the appropriateness of the term "dimension" since, as used here, dimension did not seem to correspond to physical dimensions. It seemed that what was meant by "dimension" was an aspect

of the stimulus, such as pitch, loudness of a tone, saturation, brightness of a color, etc. The questions persisted, however, emphasized by Walter Rosenblith, of whether these aspects had anything to do with dimensionality. A point was made that in statistical mechanics the dimensionality (presumably of phase space, where, for example, positions and momenta constitute "dimensions") was equally far removed from physical (geometric?) dimensionality. The question was also raised whether the aspects of a stimulus as viewed by the experimenter corresponded to the aspects as perceived by the subject. In this connection a most interesting suggestion was made by Licklider. It turns out that whereas the maximum information (from stimulus to response) on a single dimension is about 2.5 bits, possessors of absolute pitch are able to reproduce about 5 bits on the pitch dimension alone. It was suggested that for "absolute pitchers," pitch may be "two-dimensional" in the sense that one discrimination places the pitch only in a certain range, while the other discrimination identifies it as a harmonic of a precisely measured fundamental (without being able to determine which harmonic it is). The two determinations together identify the pitch. "Physically" pitch is a single "dimension" (determined by a single parameter) but "psychologically" for possessors of absolute pitch it may be two. Support for this conjecture was given by Miles S. Rogers, who cited the errors of a particular absolute pitch subject, whose errors were all either semi-tones or octaves. (The argument would be much stronger if some absolute pitchers would make mistakes by a fifth, a third, or some other harmonic interval.) Another bit of evidence (cited by Licklider) was the transmission of information in picking from a large set of chips one which matched (in color) another previously shown. The rate was 7.5–9 bits. Supposing three dimensions (saturation, hue, and brightness), this reduces to the same figure (about 2.5 bits per "dimension").

The question of how "useful" information theory has proved, or is likely to prove, again became the center of a rather heated discussion. Stated in extreme form, the position of the skeptics is that information theory is just another language for talking about statistical aspects of events. "How," they ask, "is expressing responses in bits any different from expressing them in correlations or as 'per cent'? And if there is no difference, why all the excitement? Would any one think of calling repeated conferences on 'per cent' instead of on information theory?" Similar questions are directed against the use of the term "computer" (and possibly other terms of mathematical and cybernetic terminology) with reference to human beings. Many of these remarks seem to stem from a resentment against the flood of "fashionable" terminology which gives an aura of sophistication to discourse not distinguished by either profundity or originality.

The defenders of the technological approach to psychology brought out the following arguments. Information theory is a useful tool, convenient and easy to use in quantitative psychology. Its terminology provides a bridge to the thinking of engineers. (These arguments do not satisfy those who hold that psychological theory, stated in information-theoretical terms is nothing but a "translation"). The defenders, however, stand their ground. It is maintained that information theory provides rigorously defined concepts to replace semi-respectable ones. The link to engineering is seen, as pointed out by Moncrieff H. Smith, as a source of new fruitful ideas in psychology through the medium of "engineering psychology," i.e., that field of engineering in which the human being is necessarily a component in the design.

It may be mentioned, in passing, that there is historical precedent to this argument. Physics began to make gigantic strides forward when that rather "pure" science was firmly linked to engineering. The fruits of the marriage were thermodynamics, electrodynamics, nucleonics, etc. Again the justification for information theory in psychology was related to the search for invariants, which drew the comment, "This is what I want to hear people say" from William J. McGill, a protagonist of the "skeptical" position. This point was further elab-

orated by Licklider, who related information theory to the construction of a nonparametric analysis of variance.

Donald MacKay pointed out that in the "amount of information" the psychologist has a quantity to which the concept of "flow" can be rigorously applied. Information (in that sense) can actually be considered to flow from place to place, and it makes sense to ask whether it is conserved and, if not, whether the leakages are dissipated or diverted. In response to questions about the uniqueness of the "surprise measure," on which the now classical information theory is based, MacKay conceded that it is in no way unique, except in the sense that it is the only invariant measure with respect to coding.

It was generally agreed that there may be several different measures of information or surprise, depending on the nature of the model. There was passing reference to the theory of statistical decision, which was said to be advancing more rapidly than information theory and which shows that information analysis is only one way of looking at statistical decision problems.

The next subject of discussion was context. George A. Miller pointed out the relation between this old psychological notion and the concept of redundancy in information theory. In the case of perception, for example, it might be possible to express the influence of ground on figure in terms of the covariation through which the ground is giving the perceiver some information about the figure. In the case of verbal context, the redundancy notion is most obviously appropriate and, indeed, it arose from the quantification of the notion of context. Discussion then developed around the question of how far the notion of context based on information theory can be stretched. Experimental results indicate that in the letter-by-letter guessing method of predicting the content of a sample of a language the influence of "context" (i.e., the redundancy resulting from linkages) does not stretch beyond some fifteen letters (three words). What meaning, then, can be assigned to the statement that knowledge of the subject matter of a book is an aid in guessing its content? Again, in the development of mechanical translation, determination was sought for the span of context necessary (or sufficient) to fix the one of several possible meanings of a word in a sample of discourse. It was found that two preceding words were sufficient in the great majority of cases (one could do no better with a whole sentence).

Some felt that knowledge of context should add to the redundancy "all the way down." For example, Smith remarks, "Knowing the title of the book and the author, my whole set of probabilities for various letters is going to change. . . ," to which Miller's rejoinder is, "If I tell you that this letter is *r*, and that I am thinking of *Gone with the Wind*. . ."

In general, the discussion brought out the complexity and multi-ordinality of the redundancy concept, which does not allow its mechanically identical application in all cases. Some thought was given to the results of an experiment reported by Fred Attneave III, in which subjects were asked to guess the points (in a matrix) which were on the contour of an elephant. Various groups of subjects had different amounts of advance knowledge. Some did not know what the depicted object would be; some were told it would be an elephant; some were shown the picture of the elephant for various lengths of time from .5 seconds to 16 seconds. No differences among the groups were discerned in their guessing performance. There was some discussion on the relevance of this result to the notions of context and redundancy. Alternative suggestions were made on how the experiment could be modified to bring out an improvement of guessing with increased knowledge of the pattern into which the dots are organized.

Continuing the examination of the distinction between general and "local" context, Miller suggested viewing it in terms of long-term and short-term memory, the first referring to some large locked in "set" like the English language, the second to a short span of the "stream of consciousness."

This led to the question of the invariance of information content of short-term memory storage, which certain very simple memory experiments do not seem to bear out. For

example, the span for binary digits is only one or two digits longer than for octal digits (0, 1, ··· 7), in spite of the fact that the latter contain three times as much information per digit as the former. If one, however, pursues the experiment further by training the subjects to convert binary numbers to octal (010 = 2; 111 = 7, etc.) and to store them as octals, it is found that the span becomes the same for both in *information* units (about 24 bits, i.e., 24 binary digits or 8 octal ones). We may remark in passing that here is an invariance which does not reveal itself superficially but does so if one finds the conditions for its validity, much as one finds that *in vacuo* the constancy of gravitational acceleration holds for all bodies.

The idea of language redundancy led to the question of the significance of good and bad notations and of "optimal languages." The question was raised whether languages evolve in the direction of reduced redundancy. It was pointed out that various predominant problems of communication may necessitate various language structures. The small redundancy of English "go" verbs, for example, may be a reflection of having many ways of going, "go in," "go out," "go over," "go on," "go under," etc. In a situation where all one could do was go in and out, it might be better to have two separate independent words instead of having to hunt for the right preposition.

The information theory of a visual pattern promises to provide particularly important links between engineering and psychology. Its analogues in the auditory field have actually fulfilled this promise largely through the research pertaining to efficiency of transmitting auditory signals. As is well known, the human auditory apparatus responds only to a certain range of frequencies. Therefore, it is not necessary to transmit frequencies outside of that range. This allows a limitation of bandwidth in broadcasting and in telephone. Furthermore, the exploitation of the redundancy content of spoken sounds allows a large "compression" in their transmission. The compression is possible because the redundancy has also a "psychological" meaning (besides the purely statistical one). The listener does not need to hear everything in order to hear perfectly well. The exploration of this situation in the visual area offers an interesting challenge both to the communication engineer and to the psychologist.

Edwin B. Newman opened the discussion of this subject by listing ways of defining and determining the redundancy of a picture (in television or cinema). Any such procedure stems, of course, from the way the picture is constructed, for example by scanning, as in television. This gives rise to a determination of temporal redundancy (linkage of probabilities that at a particular point there will be a white or a black square at successive intervals of time) and a spatial redundancy (the linkage of these probabilities at neighboring points at a particular time). But besides these are others, for example, symmetries, and, significantly, those of purely psychological nature, those related to *Gestalt* laws of organization. The problem of finding a fruitful approach seems related to that faced by the factor analyst in the study of mental tests. Informational problems can be handled in multiple dimensions, once these are determined, but the problem of determining what the dimensions are is a separate one (like finding the "factors" of performance), and information theory as such casts no light upon it.

Methods of determining spatial and temporal redundancies (of point-to-point type) have been developed. They yield values which are disappointingly low, allowing a compression by a factor of five, which, in order to obtain, necessitates a modulation system, such as pulse-code modulation. But such a system is wasteful of bandwidth by a factor of about five, and so the compression is of no practical importance.

This throws the attention on the other sources of redundancy. In particular contour redundancy presents a very interesting problem with inherent ramifications and connections to one of the basic areas of perceptual psychology: how does the human perceptual system handle raw information? What use is made of redundancies, inherent in gradients, contours, color data, motion, etc. in the visual field?

Attneave then raised the question of the

relative value of keeping different kinds of redundancies separate and of lumping them together. There are advantages and disadvantages in both approaches. Newman thought the values of the over-all measure of redundancy lay in that they indicated how much compression was theoretically possible, something as in the case of variance analysis, where knowledge of the total variance tells the observer how much is yet to be accounted for. It seems two strategies are predominant in the approach to the information theory of visual patterns. Some work with very complex pictures trying to arrive at some measure of their information content; others start with very simple figures, trying to work up to the complexity of pictures. As in many other situations attacked from both ends (e.g., large chaotic and simple orderly systems) exploration of the middle range, being the most difficult, lags behind.

Concept formation is a topic of psychology which has its roots deeply imbedded in the typically "Aristotelian" view of the world—the taking for granted of specific sharply defined properties inherent in all objects which allow the objects to be classified in non-ambiguous categories. In classical concept formation experiments, the subject was required to "discover" the categories, into which a set of objects could be placed (much as the philosopher of Aristotelian persuasion hoped to discover *the* categories of the world). The tendency is to become free from this tacit assumption so as to recognize the arbitrary (idiosyncratic or creative) aspect of concept *formation* as distinguished from concept *identification*. Thus, assuming that some sort of selection is involved in all concept formation, what is being selected if one starts afresh (a point emphasized by MacKay)?

It appears that the information involved in the process of concept formation must be considered relative to the subject. The finer the categories "invented" by the subject, the more information is involved (point emphasized by Newman).

Herbert M. Jenkins pursues the idea further by noting that there is more to concept formation than classification (even subjective). The perception of relations in his view is not a classification problem. It has rather to do with the generation of "rules." John R. Stroud insists, however, that any instance of use of a concept divides the world into two parts. The distinction between this and Jenkins' point of view seems to relate to the semantic (rather than to the formal, information-theoretical) content of the concepts.

Much of the discussion centered around the difficulty of coming to terms in trying to pin down the psychological process involved in concept formation. [Perhaps the application of information measures (or, indeed, of any kind of quantification) cannot be expected to be very fruitful until the term "concept formation" is sufficiently sharply defined so that various workers understand each other when they refer to different aspects of this process.]

The division of opinion seemed to crystallize around the question of whether all instances of concept formation were instances of putting things into categories (of making dichotomies into A and non-A). Those who would answer in the negative, therefore, tend to question the relevancy of studying simple classification procedures to the main problem of concept formation.

Newman invited the participants to observe the actual operation of primitive use of language—the language used by people outside of conferences. He gave a sample of a conversation between two girls on a bus.

I had a date with Jimmy last night.
Uh-Huh.
Yeah.
What did you do?
We went to the show.
Oh?
I saw George.
Yeah. George stopped to talk to me in the hall today.
D'ja have fun at the show?
Yeah.

Newman then suggests that the basic content of this sort of exchange is denotative. They consist of a sequence of "pointings" to experiences—not "judgments," which form the core of propositional logic (sequences of classifications). The ordinary "attributes" (adjectives) used in everyday speech are not primarily the simple, easily identifiable ones (red, heavy, high) but

significantly quite often rather "conglomerate sorts of things . . . [such as] . . . friendly, dangerous, aversive. . . ." These seem to be "basic in our formation of concepts." New techniques seem to be called for.

Space does not permit more than a mention of other most interesting topics discussed at this conference: Multivariate Information Transmission (the interested reader is referred to the paper of William J. McGill in *Psychometrika*, 19, June 1954); Signal Analysis; Signal Detection, and Self-organizing Automata (the last topic was discussed by Donald MacKay in *The British Journal of Psychology*, 67, 30–43, 1956, which was abstracted in *Behavioral Science*, 1, July 1956).

The spirit of the conference, however, can be well transmitted by paraphrasing some of the summarizing remarks. E. C. Poulton of Cambridge, England confessed that he received from the conference an impression completely different from his anticipation. He had expected to find in those working with information theory a "new school" (polite name for "cult"?). What he says he found was that people were quite critical of information theory, trying earnestly to determine what it will and what it will not do in psychology.

Of particular interest are the remarks of Walter Rosenblith, especially his references to the "promissory note" of information theory, which, it appears to many, is destined to provide physiology with the conceptual tools needed to weld the span from psychology to engineering. Rosenblith has a feeling that his "check is not on as short-time an outlook as . . . many people's are." He emphasized the prodigious gap between the conceptual "nervous systems" of mathematicians and engineers and the real nervous tissue of the physiologist. With reference to the relation of physiological and psychological responses, he points out that he does not even know what a "response" is.

"Now people will tell me, have I forgotten about firing and non-firing? No, . . .but we do not know how to compound responses of single units. . . . I have a hope that . . .we can get reliable organization out of reliable elements . . . but how, precisely is this done . . . not at the mathematical or conceptual level . . . but how, precisely is this done in the nervous system?"

Cautious optimism, tempered by an almost painful realization of the unfathomable complexities of behavior and of the structure of behaving organisms appeared to be the predominant tone of this conference. The transcript is recommended to everyone concerned with similar problems, not primarily for its informative content (it is too informal for that) but for what it imparts about how serious and sophisticated workers feel about these problems.

—ANATOL RAPOPORT

Information Theory in Psychology; Problems and Methods. Proceedings of a Conference, on the Estimation of Information Flow. July 5–9, 1954 at Monticello, Illinois. Henry Quastler, Editor. Glencoe, Illinois: Free Press, 1955.

The information transmission in playing tennis seems to be about 20 bits/sec., H. Quastler tells us on page 346 of the book under review. Here is a result that Shannon would hardly have anticipated from his work, but, as Deininger and Fitts put it: "Shannon's theory has provided a point of departure rather than a ready-made model." The possibilities of using information theory in psychological problems had been noticed some time ago. It is now well known that many of these problems could be handled without mentioning information specifically. Some authors have pointed out that variance and information convey essentially the same idea—at least when we are dealing with metric random variables. Yet, merely from looking at situations in the light of information theory many fruitful ideas for experimental designs are generated, and possible theoretical connections between various apparently unrelated phenomena are suggested. For it is one of the main advantages of the information measure that—being "dimensionless"—it can be applied to a great number of situations. Several contributors to Quastler's book stress the fact that this uniformity is obtained at a price: there are many aspects of a system which cannot be described or treated by the methods of

information theory; in general, it is made quite clear that information theory is of definitely limited utility.

The volume under consideration contains the proceedings of a conference on Information Theory and Psychology organized by the Bio-Systems Group of the Control Systems Laboratory at the University of Illinois, and held in July of 1954. (Some of the lectures could not be included, and, on the other hand, a few pertinent contributions not presented at the conference have been added.) Altogether, about thirty articles are collected in the book. It would have made for easier and smoother reading, if the editor (with perhaps one or two collaborators) had organized his ideas and impressions immediately after the conference into a "Handbuch" type of article on what can be done with Shannon's theories in the field of psychology. The modern trend of scientific presentation, however, seems to be against that procedure; and there are compelling reasons for it: "It was felt that information theory is a field which is in such a rapid flux that it was important to announce the results of the conference with as little delay as possible." A more integrated review article might have lost much of its value within a very short time. Quastler's presentation is thus the only adequate one, especially since numerous editorial comments are interspersed between the various articles, to help the reader fill the gaps. These editorials are always helpful and often very instructive, even when (or perhaps especially when) they deal with quite simple matters: On pp. 27–28, for example, Quastler shows by means of an elementary calculation that coding "without delay" implies only a very small loss of efficiency (with respect to the theoretical optimum) if the events considered are all equiprobable. Again, on pp. 79–82 he finds that a great number of very unlikely alternatives have little influence on the value of the information function H; this, of course, is of considerable importance; it means that in practice we can often estimate the amount of information by considering only the more probable categories of events. On the other hand, the individual papers afford a first-hand presentation of the problems as the authors saw them, an advantage which more than compensates for the frequent shifts in attention required of the reader.

The work is subdivided into three main sections: "Information Measures," "Information Theory in Psychology," and "The Utilization of Information." Of these, "Information Theory in Psychology" constitutes the bulk of the book.

Since we cannot discuss each one of the papers included, a few indications must suffice to give some idea of the general content. It is not implied that the articles we do not name are of less value than the others; in such a collection each reader is likely to find some topics of special interest. Also, this reviewer is neither a psychologist nor a specialist in information theory, and so will give a layman's impressions; some important points are likely to escape his attention.

Under "Information Measures" we may mention first the paper by Quastler and McGill (pp. 83–92) which attempts to introduce a standardized nomenclature (always a laudable purpose) and also gives a survey of multivariate information analysis, a subject developed largely by McGill and particularly important for psychology and general systems theory. The mathematical similarity, or isomorphism, between multivariate information analysis on one hand, multidimensional analysis of variance and of correlation on the other hand, is discussed in an interesting paper by McGill (pp. 56–62).

Of particular importance is the opening paper (something like a keynote address) by Lee J. Cronbach: "On the Non-Rational Application of Information Measures in Psychology," (pp. 14–26) It gives us a series of "don'ts"; we are told, in brief, that it is just as important to know what cannot be done with information theory, as it is to know the advantages which may derive from the Shannon formalism. Since the latter is based on some very specific assumptions and answers only certain types of questions, a psychologist using the theory should make sure that the conditions under which he operates are similar to those of Shannon at least in some basic respects.

A feature which is of little importance in

technical communication theory but constitutes a real problem in psychology concerns the sampling distribution of information functions. We use a set of probabilities in order to calculate the uncertainty function, and these probabilities, in practice, must be estimated from the relative frequencies with which certain events occur in a finite number of trials. As is well known, our estimates will differ from the "true values" because sample statistics do not agree with those of the parent population from which the sample is drawn. G. A. Miller and W. G. Madow have derived an approximate formula to calculate the sampling bias of estimated information functions; their work is reported on pp. 95–100. M. S. Rogers and B. F. Green calculate exactly the first and second moment for the sampling distribution of H when all possible outcomes of the event considered have equal probabilities. Finally, L. G. Augenstine, uses the ILLIAC computer to generate artificially a great number of probability sets such as might be obtained by taking many samples from the same parent population, and thus constructs the sampling distribution "experimentally." Unfortunately, work was still in progress at the time of printing, and only one or two preliminary results could be offered.

In the final paper of this section (pp. 124–139) H. Quastler discusses various experimental methods and ingenious techniques to obtain approximate values or to establish bounds for the amounts of equivocation and transmitted information in a human channel.

Perhaps the most significant article in the principal section is H. Quastler's "Information Theory Terms and Their Psychological Correlates" (pp. 143–171). The main points which he discusses are treated subsequently and at greater length by other contributors. Without going into details we should mention a few topics whose importance emerges from Quastler's presentation:

In many situations of interest in psychological research human subjects are considered as channels transmitting information. Shannon's theory may be useful in giving a description of what these channels can achieve. As to how they work, results obtained from applying information theory are at best useful hints; the neurophysiological approach appears more promising in this respect. In technical communication systems properties of channels may be deduced from their known structural characteristics; in psychology, those properties (basically a matrix of input-output conditional probabilities) can, at present, only be inferred from observations on how the channel functions.

"Amount of information" always refers to a carefully defined situation; we must have a well-determined universe of discourse. The question then remains: what are the units (elementary symbols in Shannon's sense) which human beings use to process information, and how many bits are they "worth" (what is their informational value)? John M. Stroud (pp. 174–205) presents evidence that conscious time is quantized. As yet he is unable, however, to state definitely how much information is handled in one of these psychological time quanta ("moments"), or, rather, in one of the hypothetical higher order moments (cf. Stroud, p. 202, also Quastler, pp. 347–349). E. L. Fritz and G. W. Grier, on the other hand, find it convenient to use "pragmatic units" which they call information elements (pp. 232–243, in part. p. 236). These are defined operationally by breaking down messages into sequences of words such that each sequence or element would be just sufficient to alter the subject's response. That this analysis could be carried out with high accuracy is a remarkable fact, and the authors deserve credit for having devised this method.

The number of categories or classifications which are relevant to a given message, or, in other words, the various possible answers that could be given in reply to a certain question, always plays an important role in assessing the amount of information received and transmitted by human subjects. A. Rapoport likes to illustrate this point as follows: The Japanese have two ways of saying "grass is green;" one of them in answer to the question "what color is grass?", the second in answer to "what is green?"

Such considerations bring us into the do-

main of semantic information, some elements of which, of course, appear also in the experiments by Fritz and Grier just mentioned. Semantic information as such, however, is not one of the main subject matters in Quastler's book.

Since the information function is based on a set of probabilities, it is clear that the problem of objective vs. subjective probability will be of great importance in psychological experiments. In particular, whenever information received is utilized subsequently in making a decision (a very common occurrence indeed), the subject is likely to evaluate whatever information he can get in terms of his own subjective probabilities. These questions are reviewed and discussed in a paper by Hake (pp. 257–277); it appears that human subjects are quite accurate when predicting the occurrence of a next event in sequential situations. It may be interesting to notice, since we are speaking of probabilities, that the notion of "personal probability" plays a fundamental role in L. J. Savage's recent book on statistics.[8]

Finally, Quastler discusses information theory from the point of view of organization, i.e., interrelations between the parts of a system in terms of information transfer. Multivariate information analysis thus becomes a valuable tool for the general theory of systems.

Much of the discussion in the remaining papers of this section is concerned with the optimum performance of subjects in transmitting information, that is, the channel capacity of human beings.

For selection of events pertaining to only one classification there may exist a common maximum value of performance (measured in bits of information transmitted), largely independent of the kind of stimulation. This could then be called a channel capacity, as long as the input is limited to that one variable. H. W. Hake, however, doubts that this will hold for any kind of "unidimensional" inputs, whereas Quastler is more optimistic on this point (pp. 248–253, 254). This controversial matter has been brought up again quite recently in a very interesting paper by G. A. Miller.[9] It is clear that if the experimental results collected by Miller were expressed in terms of variance they would not show such an impressing uniformity over various inputs. However whether the repeated occurrence of the "magical number seven" is more than a coincidence, cannot be decided by information theory alone. There is no doubt, incidentally, that Miller's paper would have to be included in a future edition of the book we are reviewing. It throws light on two other points discussed by Quastler in connection with organization, namely (*a*) the fact that human beings handle information by collecting elementary units into "chunks" of higher informational value (unitization), and (*b*) the span of influence of relations or units. This latter subject is closely allied to redundancy; we may, therefore, mention here a paper by Virginia L. Senders and J. Cohen which describes some experiments on sequential dependencies in instrument reading performance (pp. 282–289). Adelson, Muckler, and Williams, on the other hand, present what amounts to a simple theory of learning based on the concept that the subject's uncertainty is reduced in a constant proportion at each trial (pp. 291–299).

The tone of the book is generally optimistic, in spite of the fact that so many variable factors influence and determine the functioning of human channels. The limitations of information theory are clearly recognized, but also its power (if properly applied). Thus Leonard (p. 313) writes: "The fact that this type of work was *not* carried out earlier although the tools for it probably existed is some indication of the extent of our reorientation."

Stimulus discriminability (in space and time) and complexity are clearly two of the factors which influence channel capacity (Leonard, pp. 306–314). However, the *combined* characteristics of both stimulus and response, or the different ways in which a subject may be asked to respond to a given stimulus have a decisive effect on rate and

[8] L. J. Savage, *Foundations of Statistics*, New York: John Wiley & Sons, 1954.

[9] G. A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review*, 63, 81–97, 1956.

accuracy of performance. Deininger and Fitts call it "stimulus-response compatibility" and illustrate its meaning by a series of very neat experiments (pp. 316–341). "It is not possible to overestimate the importance of this concept; in studying compatibility, one approaches the most intimate properties of a channel" (Quastler, p. 315).

As these inquiries on the capacity of human channels progress, it becomes evident that many features must be added to the original kernel of Shannon's information theory. It is not too surprising, therefore, that neither of the two formal definitions of channel capacity provides an adequate means for measuring it (Quastler, pp. 341 ff.) The wealth and variety of conditions under which a human channel can handle information is not nearly approached by any of the technical communication systems.

Other matters, certainly deserving attention but which can only be mentioned here are: the linear relationship frequently found between reaction times and amount of information transmitted (Bricker, pp. 350–359), Quastler's and Virginia Sender's remarks on informational aspects of Gestalt in visual perception (pp. 300–302), and Osgood's paper on "Fidelity and Reliability" (pp. 374–384) "which goes beyond the informational analysis of channel performance."

The last main section of the book includes three papers on the theory of signal detection, where the "utilization of information" has been most extensively investigated. The final remarks in Tanner's article (pp. 403–414) should be particularly important for the psychologist.

In conclusion, Quastler's book succeeds admirably in setting the relations between information theory and psychology into their proper perspective. There are several printing errors throughout the volume (though not enough to cause serious equivocation), and the legends of figures and graphs are sometimes difficult to read. These minor inconveniences do not detract from the value of the book, but they might be eliminated in subsequent editions. Psychologists who want to read Quastler's book should take into account that considerable knowledge on the basic concept of information theory is assumed throughout.

—ERNESTO TRUCCO

Toward a Unified Theory of Human Behavior. Roy R. Grinker, Editor. Other contributors: Karl Deutsch, Alfred E. Emerson, George L. Engel, Lawrence K. Frank, Jules Henry, Florence Kluckhohn, Howard Liddell, Charles Morris, Talcott Parsons, David Rapaport, Anatol Rapoport, Jurgen Ruesch, David Shakow, Alfonso Shimbel, John P. Spiegel, Laura Thompson, James E. P. Toman, and Paul A. Weiss. New York: Basic Books, 1956.

A new trend is developing in behavioral science theory, represented by three publications of similar title as well as thought. First there was Parsons and Shils's *Toward a General Theory of Action*, then an article, "Toward a General Theory for the Behavioral Sciences," and now a new book, TOWARD A UNIFIED THEORY OF HUMAN BEHAVIOR. Some say this is an untoward movement; some say not. However that may be, the facts are that it is so delicately suspended between importance and grandiosity that it has already been impaled upon a partially deserved parody entitled, "Sidesteps Toward a Nonspecial Theory."[10]

In 1950 Roy Grinker and Jurgen Ruesch decided to approach the problem of developing a unified theory of human nature. After discovering a number of others who shared their enthusiasm for the project, they arranged for the first of a series of biannual weekend conferences, which was held in October, 1951. The gamut of behavioral disciplines has since then had distinguished representation in these conferences, including biologists, physicians, and social scientists of diverse backgrounds.

The participants were not a random sample of the whole range of behavioral scientists, but they were representative enough of the sciences of man to indicate dramatically the great diversity of approaches to the study of human nature which can be found in the different behavioral sciences. In one of his summaries (pp.

[10] Borgatta, Edgar F., *Psychological Review*, 61, 343–352, 1954.

135–136) Grinker commented on the four categories into which he believed the speakers fell—those who emphasized the observer and his location in relation to the phenomena observed; those concerned with describing organized systems of behavior; those who employed conceptual models; and finally those who concentrated on transformations or processes in systems from one state to another and transactions between systems. In addition, the interests of each participant centered around special aspects of behavior—communication, group dynamics, value theory, or evolution, for example. With so many dimensions available for deviation, no two participants shared identical viewpoints.

The book is a condensed and edited report of the ebb and flow of the first four conferences (out of ten so far held). The excitement of the lively back-and-forth play of ideas appears repeatedly, and at some points the reader views the birth of a new conceptualization. But, lacking gesture and intonation, the book differs from real life the same way the written script differs from a play. Despite days of editing by Grinker and his assistant, Helen MacGill Hughes, the conversation is often disjointed and lacking in the subtlety it undoubtedly had originally. At some places the text nears incomprehensibility, perhaps because of the extensive cuts which were made. Editing such protocols presents almost insuperable problems, but the value of them is often such that it is worth trying to overcome them. (An index would have been worthwhile also, but since none exists, page references are liberally sprinkled through this review.)

Publications like this reveal in a hot, harsh light the participants' thought processes, the interchanges occasionally resembling a sort of intellectual strip paper. The proceedings' largely unrehearsed character heightens the reader's enjoyment, particularly in the pleasant observations of outstanding competence, as for example, certain statements of Deutsch, Weiss, Rapoport, and Grinker. Of course, it is natural that large and important areas of behavior theory were brushed over lightly or entirely neglected—problems of structure or the field of perception, for instance—however, this is to be expected in any such limited document and many of these issues were undoubtedly raised in later conferences. Despite the magnitude of the task faced by the group, the minutes contain evidences of real increases in understanding, sharpening of ideas, and implicit and explicit extension of agreement among the participants. (Compare, for instance, pp. 37 and 38 with pp. 340–350.)

The participants in the conferences recognized the classic difficulties of communication, terminology, and semantics in the slowly interacting "group mind" which is inherent in any body that comes together only biannually in order to carry out such intricate tasks. They approached their work realizing that it would go slowly, and with appropriate humility stressing that their goals were "a first approximation to a scheme which will enable us to represent physical, psychological, and social events within one system of denotation" (p. xi). And the editor adds that "the members of the conference were committed with great zeal and seriousness of purpose toward a program of developing a unified theory of human behavior. Yet all of us implicitly knew that this could in reality not be obtained by us in our time and that we would be satisfied with a little progress toward the ultimate goal. It was amazing that through the years the group continued to be overtly optimistic, each member continually striving to understand various aspects of human behavior discussed by the others" (p. 366).

Convergence Upon Systems Theory

Despite the divergence of viewpoints among the discussants, from the very first presentation, there was continuing emphasis on the notion of "systems," as an integrating concept. Apparently this emphasis was the result of independent thinking by several members of the group. In this they appear to share a *Zeitgeist*, because other would-be integrators (the Macy cybernetics symposia, the group allied with Parsons and Shils, Operations Research teams, and so on) seem to have arrived independently at similar conclusions. But, like others before and after them, the participants in this conference

found it hard to agree on exactly what a system is. The following three notions of systems often were confused and never were precisely distinguished:

a. Conceptual systems—formalizations in the sense ordinarily employed in mathematics.

b. "Real" systems, living or nonliving—objects in physical space-time which are observed and measured ordinarily by methods and procedures common to the natural sciences.

c. Abstracted systems—either relationships of various sorts or classes of behaviors or relationships which can be identified in, or exist between, "real" systems.

There were some exceptions, but in general the participating behavioral scientists with primarily mathematical backgrounds tended to think in terms of conceptual systems; those with biological backgrounds emphasized "real" systems; and those primarily experienced in the study of broad social processes dealt with abstracted systems.

The differences between scientists favoring category *a* and those favoring category *b* are well illustrated in the following interchange (p. 328):

RUESCH: Previously I defined culture as the cumulative body of knowledge of the past, contained in memories and assumptions of people who express this knowledge in definite ways. The social system is the actual habitual network of communication between people. If you use the analogy of the telephone line, it corresponds to actual calls made. The society is the network—the whole telephone network. Do you agree with these definitions?

PARSONS: No, not quite. In the limiting conception a society is composed of human individuals, organisms; but a social system is not, and for a very important reason, namely, that the unit of a partial social system is a role and not the individual.

RAPOPORT: The monarch is not an individual, but is a site into which different individuals step. Is that your unit of the social system?

PARSONS: Yes. A social system is a behavioral system. It is an organized set of behaviors of persons interacting with each other: a pattern of roles. The roles are the units of a social system. We say, "John Jones is Mary Jones' husband." He is the same person who is the mail carrier, but when we are talking about the mail carrier we are abstracting from his marriage relationship. So the mail carrier is not a person, just a role. On the other hand, the society is an aggregate of social subsystems, and as a limiting case it is that social system which comprises all the roles of all the individuals who participate.

To Ruesch the social system seemed a "real" system in space-time; observable and presumably measurable by techniques like those of physical science. To Parsons the system is abstracted from this, being the set of relationships which represent the organization. These relationships are conceived of as abstractions and measured in behaviors of subsystems in the society. To Parsons the important units are classes of input-output relationships of subsystems rather than the subsystems themselves.

Grinker accurately described this fundamental, but not irresolvable, divergence when he said: "Parsons stated that there is a distinction between a frame of reference and a class of systems. Action is not concerned with the internal structure of processes of the organism, but is concerned with the organism as a unit in a set of relationships and the other terms of that relationship, which he calls situation. From this point of view the system is a system of relationships in action, it is neither a physical organism nor an object of physical perception. On the other hand, some of us consider that the foci or systems which are identified in a living field must be considered as being derived through evolution, differentiation and growth from earlier and simpler forms and functions and that within these systems there are capacities for specializations and gradients. Sets of relationships among dimensions constitute a high level of generalization that can be more easily understood if the physical properties of its component parts and their origins and ontogenic properties are known" (p. 371).

The following generalizations were made in the conferences concerning some or all behaving systems:

Grinker pointed out that, unlike nonliving closed systems, in which there is degradation of energy, living systems are open, and their throughputs can prevent an increase in total entropy. Furthermore, rapid transformation of energy among the subunits of living systems is possible, as he indicated. In these

units also energy can be stored or bound. There are boundaries to these systems, and there is exchange of substance across these boundaries in both directions. These systems interact with their environments by circular mechanisms of feedback. They maintain states of activity which are constant within a certain range of variation. The most intense activities of such systems are most internal or central in them. Built-in tensions or drives activate such organisms. This sort of activation occurs even before sensation concerning the external environment develops.

Spiegel contended that it is more desirable to conceive of systems as interdependent foci rather than hierarchical organizations. They are involved in multiple reactions—transactions with several similar foci, which are more complex processes than simple interactions. He found three sorts of patterns of process in systems: (*a*) the constitutional determinants, which concern the structure of a system, and which can be observed along spatial dimensions; (*b*) integrative determinants, which concern function of the system, and can be observed along temporal dimensions; and (*c*) system determinants, which deal with the role or purpose of the system in the larger system of which it is a part, and which can be observed by examining the transactions and interrelations of the system—what might be called its external relations.

As principal foci of transactions Spiegel listed soma, psyche, group, society, culture, and universe, each involved in multiple transactions with the others. In separating society and culture, and soma and psyche, he appears at times to fall into confusions which are at least partially related to the mentalistic or subjectivist problem that will be discussed later.

Also Spiegel indicated that there are defenses which prevent disintegration of systems, but he mentioned only one sort out of several—encapsulation or isolation. He also stated that stresses or strains can impinge on systems from outside.

To these observations Shakow, speaking at least of individual human beings as systems, indicated that there are mechanisms of delay of gratification or of substitute gratifications of needs. He categorized under these mechanisms several different mechanisms of defense. He also stressed that such systems are capable of learning.

Growth and development were emphasized by Engel as capabilities of such systems. He indicated that they adapt to the environment within genetically limited capabilities, and he pointed to their ability to reproduce themselves. At one point in the discussion he took issue with Spiegel's nonhierarchical organization of interacting foci, and apparently came closer to the view expressed by Grinker—that there is a hierarchy of behaving systems.

Speaking primarily from a communication theory framework, Deutsch later added to the above observations about systems that they contain certain critical components in which a small change will lead to a bigger change in behavior of the whole system than a comparably small change in another unit. He also referred to nodes, which are units in systems that have a particularly high statistical frequency of transaction. He likened these to the intersection of roads or to a switchboard in operation. Further he added that systems have memory pools and that they are equipped also with decision points with varying degrees of freedom of decision. He indicated also that at least some systems contain high-order feedbacks, which abstract information either from the memory or from current decisions or current inputs and give a higher-order of response to them. Some of these responses are sometimes called conscious processes.

And to these other observations Ruesch added that systems can have goal-seeking devices, and also they can have goal-changing devices.

The central plot of these discussions revolves around this major problem of systems and their characteristics, but as Grinker recognized in his final summary, there were secondary plots which concerned the review of what the group gradually came to feel were central questions—particularly communication of information and equilibrium or homeostatic control.

Information Theory

The progress of the discussions indicates that throughout the sessions interest increased in the fundamental concepts of information or communication theory and the related field of cybernetics as applied to the understanding of behaving systems. There was much discussion back and forth about the fundamental notions of this theory, such as transmitter, receiver, channel, message, noise, and feedback.

Though Parsons expressed grave doubts (p. 52) concerning the applicability of Wiener's theory to the phenomena of social science, he appeared to have been in a minority. Ruesch made an effective interpretation, complete with diagram, of interpersonal communication in terms of the concepts of information theory, going into details on various aspects of it such as the coding process. Also he introduced into the discussion the notion of metacommunication defined as all the interpretative devices which are used in the exchange of messages—"explicit instructions, implicit instructions contained in role assumption, reference to context of the situation, rules, sequences, patterning," and so forth (p. 47). He went so far as to state that the concept of noise has no place in the theory of communication, because anything which enters the behaving organism will shape the content of the communication and will not, as in mechanical systems, simply reduce its intelligibility. On this issue there was disagreement, because Deutsch (p. 285) made a clear distinction between noise and steady biases in the communications of behaving systems. There were further discussions of the relation of noise to entropy and the signal-to-noise ratio. In a stimulating discussion Deutsch (pp. 349, 354–357) indicated that the boundary, which concept appeared so important in systems theory, could be measured in terms of the rate, richness, and a number of other characteristics of transmission of information across it. He was interested in this in terms of large social and political systems, but his thinking is obviously applicable to other types of systems. Altogether he listed twelve empirical dimensions related to information theory, which could be employed in studying boundaries.

Deutsch also discussed other behavioral processes in information theory terms, such as predictions, restructuring, learning, "delegation," "secession," and "integration," the last three concepts being related to organization of systems. He even dealt with the concept of growth in information theoretical terms.

Included also was an insightful discussion of how abstraction and condensation are essential to permit maximal storage of contents in the memories of behaving systems. He pointed out that (p. 260) "in order to learn and remember and teach, both the individual and the society have to throw away fantastic amounts of detail. You remember by flattening and sharpening images, patterns, rumors or the like. If this is so, then there comes a stage where we have thrown away so much detail, and simplified so much, that we can no longer reconstruct the earlier ensemble from which the latter patterns were generated. Irreversibility thus would occur where generative combinatorial universes have become critically modified or depleted by selection. . . . Social development, in particular, always involves stages of great simplification. It must be irreversible because very soon the information needed for retracing the intermediate steps is lost. Irreversibility, in other words, is the opposite of novelty, and novelty is the opposite of the economy of recording facilities."

He also made a final important point in distinguishing between energy-linked systems and trigger-linked systems, most of which are systems linked together by the transmissions of information. He pointed out that when the amplification ratio of the trigger mechanism is very high (when the proportion of the output to the input is great), or when the output has much unpredictability or many degrees of freedom in terms of the input, then the systems seem to have great autonomy. He therefore suggested quantitative ways to distinguish among the more complex behaving systems which appear to have a great deal of autonomy, and those which have less.

Grinker added to the discussion of in-

formational concepts a valuable, although relatively vaguely formulated, statement of similarity between energy flows and information flows, recognizing formal identities between such energy phenomena as irritability and motility and such information phenomena as tension and consciousness, or between the energy phenomena involved in growth and the information transmissions involved in learning, or between the propagation of energy and propagation of information in memory and recall (p. 138). And on a related topic Toman (p. 360) argued that fascination with new developments in information theory should not cause us to neglect the energy transmissions which must exist between any systems that communicate.

A different, and as yet generally unexploited but highly significant aspect of information theory, was raised by Weiss. He distinguished between energy and form, structure, or organization. Then he asked (p. 114) "Is organization an all-or-none property, or amenable to gradations and variations? Can we arrive at any quantitative scale by which we can rate organization?" Emerson agreed with him (p. 158) as to the need of discovering the dimensions of structure which exists in the "real" world, and developing measures of it.

Among the most creative inferences from the conceptual framework of information theory came in Emerson's suggestion (pp. 216–217) that "The symbol as a means of communication and continuity presents some suggestive and stimulating comparisons in the dynamics of cultural evolution in contrast to the gene in the dynamics and mechanisms of biological evolution." He suggested that this comparison be carefully investigated down to small details, a suggestion which in more recent years experimenters are beginning to follow up by studying the spread of culture in terms of what is known about the genetic spread of biological heredity. Throughout the discussions were provocative comments about inheritance and evolution in information theoretical phraseology—e.g., Emerson's discussion (p. 245) of "regressive evolution;" Toman's suggestion that living organisms transmit through their genes to succeeding generations information concerning their environments; and a consequent notion that the sexual genetic processes are essentially processes of "pooling of information" (p. 251). By such mechanisms members of a given species are enabled to maintain forms of homeostatic equilibrium that would not otherwise be possible.

Homeostasis

The discussion of homeostasis was closely related to that of evolution. It has been repeatedly affirmed that natural selection produces more complex individuals, but Emerson pointed out that it also leads to an increased efficiency of specialized functions within behaving systems, consequently to better homeostasis in the environment, and thus to longer survival. He stated (p. 178), in terms reminiscent of the probability measures of information theory, that the effect of natural selection is to make the improbable become probable, which it does by increasing the complexity of organisms. Emerson views homeostasis philosophically, as a principle of behavior with the broadest implications. While outlining six specific elements of homeostatic phenomena (p. 235 ff.), Rapoport quoted a good summary of the principle from a statement made in 1900 by Charles Richet, "The living being is stable. It must be so in order not to be destroyed, dissolved or disintegrated by the colossal forces, often adverse, which surround it. By an apparent contradiction it maintains its stability only if it is excitable and capable of modifying itself according to external stimuli and adjusting its response to the stimulation. In a sense, it is the necessary condition for the true stability of the organism."

Deutsch suggested three potential empirical measures which would indicate whether process A is "more homeostatic" than process B:

(*a*) If the average deviations of A from the level maintained are smaller.

(*b*) If the range of loads which will not throw it out of gear is larger.

(*c*) If learning proceeds so that one range of loads can be dealt with better

although the organism is more easily disturbed by other ranges of loads (p. 243).

There was lively discussion of other interesting problems relating to homeostasis: the role of negative feedback and amplification in homeostasis; what it means for a larger system to take over responsibility of homeostasis for one of its subsystems (as when a family cares for a sick member); and the relationship of homeostasis to apparently purposeful behavior.

Problems of Method

Certain methodological problems were discussed in detail, such as what constitutes a theory; the difference between ordinary theory and "theories of theories" like growth theory, homeostasis, and information theory, which deal with general functions and do not have content specificity (p. 157); the role of analogies and formal identities in general theory construction; the number of observations requisite for generalization; and the place of mathematical models in behavior theory. Good examples of such models were discussed in detail by Rapoport, Toman, and others.

An issue which repeatedly came up was the problem of the observer and his position in relation to the phenomena—the difference between the external and internal view of a behaving system. Thompson commented that following the publication of Einstein's early work, Picasso and his successors in modern art have been concerned with a similar problem, that of seeing an event from within and without simultaneously. Parsons observed that modern mathematical methods like the calculus give us techniques for looking at complementary things at the same time.

Issues for the Future

Among all these comments there were many unresolved points, some because the problems were not well stated and others because—even though they were well stated—there appeared to be recurrent causes for explicit or implicit confusion. Perhaps the most pervasive difficulty might be referred to as the "mentalistic" confusion. The question was never clearly asked or answered as to how a mental phenomenon can be differentiated from a physical one. It was recognized that many behavioral scientists, including apparently some of the participants, felt that mentation was more than an organic function, but in what way was never specified. That "psychic energy" differed from physical energy was granted, but in what manner remained vague. Consciousness was equated to "high-order feedback circuits," but what this meant was not made explicit. There was confusion about the difference between the information in an interpersonal communication and its meaning. There was misunderstanding as to how game theory or information theory could refer to decisions being made without a person being involved.

The result of this pervasive and many-faceted confusion was that many of the rubrics and categories of conceptualization appeared not to be entirely self-consistent.

It was apparent in efforts by Shakow (p. 27), Spiegel (p. 169), and Grinker (pp. 332–333) that the classical formulation of Freudian psychoanalysis does not fit in smoothly with the systems-theory approach of the group. Just why this is so was not clarified, although perhaps it is because Freudian thinking is based primarily on a hydrodynamic model and systems theory more on a field model.

Another interesting unresolved question was whether life is more than mere homeostasis or tension-reduction. The role of symbolic goals and purpose in life was considered but not completely clarified.

In a number of discussions hierarchy of systems was confused with levels of abstractions and the various fundamental definitions of systems which we have discussed earlier. Since the various fundamental definitions were never clearly distinguished this had a constant muddying influence on the discussion. Particularly the social scientists seemed to feel that the symbolic and cognitive behavior of human beings and human individual groups in society was so much more complex than that of lower species and quantitatively so different that it was essentially a qualitative difference. To the biologist this did not appear to be such

a difficult problem, although apparently some of the biologists were on the side of the social scientists and vice versa. The differences primarily emphasized between behavior of social systems and of systems at the individual and lower levels were that societies showed (1) greater use of symbols and transactions at a distance; (2) more locomotion of component units; (3) less integration, or as Weiss put it (p. 123) "totalitarianism;" and (4) more rapid reallocation of resources in the society. Until these differences were accounted for it appeared that the notion of hierarchical levels of behaving systems would not be generally and unquestionably accepted in the group.

A significant issue almost completely neglected in the first four conferences recorded in the book is the problem of utilities and values. Florence Kluckhohn reported her comparative study of values in different cultures but the relationship of this to the general problem of the group was not explicitly spelled out. Charles Morris pointed out (p. 277) that values may differ between systems and subsystems, progress for one not necessarily being progress for the other. But these brief comments were the sum of comments on these topics.

Perhaps the pervasive tenor of the book can be best summarized in the following quotation from Thompson, (p. 72) who indicates the characteristics not only of a good theory of culture but also of a good general theory of behavior:

"1. Human culture had to be recognized explicitly as the product and process of human organisms and therefore as a biologically-based phenomenon.

"2. Having recognized the biological base of culture, it was necessary to have an adequate theory of organism on which to build a biologically-based theory of culture. In other words, the science of biology had to develop beyond the bio-physical, biochemical phase to the functional integrative phase.

"3. This theory of organism had to account not only for the activities of single organisms but also for the activities of groups or communities of organisms in environmental setting: e.g., it had to account for natural ecological arrangements and communities. In other words, the science of integrative ecology had to develop beyond the phase of general description to the phase of explanatory description.

"4. An adequate theory of personality had to be evolved which would account for the growth, development, organization and integration of personality in total relevant environmental setting and under changing pressures from within the human organism and outside it; which would account for the decisive role of culture in the development of personality; and which would account for the creative, purposive role of personality in the building, integration and regeneration of cultures.

"5. The science of linguistics had to develop to the point where the interrelationship between linguistic structure and culture structure emerged.

"6. An integrative, interdisciplinary methodology had to be developed which was adequate to describe and analyze not merely social behavior but a whole culture-in-environment in spatial and temporal perspective, including its ecologic, somatic, sociologic, psychologic and symbolic dimensions.

"7. And finally, human communities which had been relatively isolated through millennia under extreme environmental pressures had to be analyzed multi-dimensionally so that the processes of culture creation, culture growth, culture adaptation and integration with the environment, culture crisis, culture reintegration and culture disintegration could be understood in geographic setting and in long-range historical perspective."

And of course she recognized, as did most participants in this conference and as do most behavioral scientists, that this is an ultimate and that in the meantime science can effectively continue simultaneously at all the stages she mentioned.

—JAMES G. MILLER

J. T. Bonner's *Cells and Societies*[1]

The sameness of living things and their wonderful variability are the subjects of this book in which the different means used by animal societies, cell colonies, and simple and complex individual organisms to achieve the ends of feeding, reproduction, and communication are examined and compared. The author believes that all living things have a certain sameness and that some of the basic principles of life can be learned by this study of how each of these, as a unit, performs these necessary functions. Methods used by living organisms in securing and utilizing oxygen, digesting food, eliminating wastes, and reproducing are also examined. Because of the simplicity of single-celled and colonial organisms the author says that the fundamental meaning of living activities can be understood through studying them and that the functions of complex multicellular plants and animals will be more easily understood.

Societies of howling monkeys, fur seals, red deer, and beaver are described first, followed by the social organizations of social insects such as bees, army ants, and termites. The activities of hydroids and of cell colonies such as obelia, zoothamnium, and volvox are then taken up. The single-celled amoebae which usually live independently but collect into cell aggregates with communal and integrated properties when the food supply is low are among the more interesting of these colonies. Finally free-living single-celled organisms which differentiate their functions on a molecular level are described.

Societies have evolved independently in a number of different lines and have different characteristics while securing similar advantages for their members. The society of the howling monkeys achieves the goal of reproducing itself by continued living together so that the female is bound always to be in the presence of males and available for sexual relations, while the fur seals live and feed as solitary beings except during the short Northern summer when they instinctively return to the rocky islands on which mating and the birth of young occur. The female seals have evolved in such a way that one horn of their bicornate uterus can be fertilized while the other is recovering from the birth of a pup. Fortunately, also, the gestation period in the seal allows the young to be born about a year after fertilization and thus come into the world at the annual rendezvous with the rest of the society.

Reproduction among social insects is carried on by specialized members only and among colonial cells. There may be complex sorts of cell divisions while the new cells remain as a part of the colony or, alternatively, may become free-swimming individuals and form a new colony elsewhere.

Communication among the higher animals frequently involves distinctive sounds which all members of the group recognize. The howling monkeys make use of a "language" of nine sounds including the typical howl of the males which indicates some sort of disturbance, a sort of "cluck" which is used to indicate that a pathway has been found, and a sound made by young monkeys when they fall to the ground. They also use a gesture language.

The beavers use a sharp slap of the tail and a splashing dive to give the danger signal while the red deer sentinel has a warning "bark." Communication and coordination among the social insects are achieved through their chemical senses of taste and smell, through sounds such as wing vibration, through exchange of food, and in such elaborate ways as the complex "dance" by which bees describe the location of a distant food supply.

Coordination in cell colonies may be achieved through a contractile thread which carries an impulse that causes all cells to contract or it may involve a chemical

[1] *Cells and Societies* by J. T. Bonner. Princeton: Princeton University Press, 1955. Pp. 234, $4.50.

reaction such as governs the aggregation of amoebae. Communication and integration in more complex animals involve nervous systems and hormonal systems. In the words of the author: "... all these similarities and diversities are variations within strict confines, the same strict confines that affected mammal and insect societies, the need of a living unit to perform certain specific living activities." (p. 114)

The chapters which follow the analysis of the activities of these various living units examine "what is really meant by feeding, reproduction and coordination" (p. 124). The need for energy and the importance of enzymes is traced. Photosynthesis and chemosynthesis among plants and bacteria are discussed and the dependence of animals upon plants for synthesis of edible materials is explained. The processes of digestion and elimination of wastes, of gas exchange, and the securing and utilization of oxygen are compared at the levels of colonies, cells, and complex organisms. Reproduction and coordination are also discussed.

In all of these the emphasis is on the variety of ways in which the basic requirements of the organism are met. For example, the blood of a species may be blue (insects), green (worms), or red (vertebrates and some insects), but blood always contains a complex protein involving a metal atom which can combine efficiently with oxygen. The green and red colors are caused by different compounds of iron; the blue is a copper compound. The result is the same: oxygen is carried into the body for use in its many chemical reactions.

The final chapter briefly points out that human society is similar in function to the others that were discussed, since man is a living organism which must carry out the same activities of feeding, reproducing, and communicating as do other living things no matter how simple or complex. The ways in which human societies differ from the others are conditioned by the ways in which people differ from the members of other societies: they have less instinctive behavior and more learned and culturally determined activity. They have much more complex nervous systems and thus more complicated modes of response. However, although man may be variable and subtle, "He too is made of protoplasm, and no matter how elaborate is his brain and his social structure, he can never escape from the rules laid down by his body substance."

This book is written in an easy and popular style in which technical terms are avoided or simplified. It is evidently aimed at a nonprofessional audience. At times the outline is obscure so that it is sometimes difficult to see the intended relationships among the parts.

The fundamental point is made that the similarities observed on the different levels from cell to society are not superficial analogies but are truly comparable in terms of the need of living things to meet conditions imposed upon them by their structure. It is the author's belief that scientific understanding at the levels of cell and animal societies contributes to human behavioral science. (*JM*)

Let's not be too quick to blame the human race for everything. A great many species of animals became extinct before man ever appeared on earth.

—Will Cuppy, *How to Become Extinct*

W. W. GRINGS' *Laboratory Instrumentation in Psychology*[1]

The preparation of an instrumentation text in psychology is a more formidable project than for most other disciplines. The proper object of study for experimental psychologists is undefined and perhaps undefinable. In consequence their need for tools of the trade may encompass instruments appropriate to any of the life sciences and more than a few from fields of inanimate study.

A text of this sort must choose between being a catalog of current equipment and manufacturers, a text in applied engineering, an experimental psychologist's *vade mecum*, and a student introduction to principles of instruments in common use by psychologists. The latter choice seems to be the one the author has generally made, though occasional lapses are evident. The book is expressly intended for the use of graduate students and professional people who have an interest in a new area of research but lack the technical information necessary to get started. Whether so intended or not, it would also serve well as a textbook in the courses in psychological instrumentation lately appearing at most universities.

The book has five main topics: a chapter each is devoted to recording systems and timing and counting mechanisms, three chapters are devoted to sensory stimulation and response measurement with vision and audition each rating a separate chapter, another chapter is devoted to human learning and perception, and a short concluding chapter discusses bioelectricity. In addition, there is a short reference to sources of apparatus consisting mostly of titles of publications devoted to principles and design of instruments. The Grings text is fairly comprehensive for the limits of scope set. Its chief advantage lies in being the first text in psychological instrumentation since Wendt's monograph appeared in 1938.

Considering the range of effort the author has undertaken it is not surprising that many readers will notice significant omissions. Very few specific solutions to likely apparatus problems are actually offered. For example, the discussion of bioelectricity offers no source of probe electrodes either for stimulating or recording nor any recipe for their manufacture, although sketches and recipes are provided for skin electrodes and the associated paste electrolyte.

Some of the text devotes more than the necessary space to instruments which are of more historical than practical interest. This is particularly true of the sections on timing devices and on recorders. Pendulum timers, Hipp chronoscopes, and muscle levers writing on smoked paper have been almost entirely supplanted in the modern psychological laboratory by electronic timers and counters, synchronous motors, and the oscilloscope. Although pen recorders writing in ink on paper are extensively used (and the discussion of these is quite good) they are not of the general utility one may find in a good oscilloscope and an associated kymograph camera. The immediate and obvious advantage of a permanent record has frequently led experimenters to overlook the many great advantages of an oscilloscope and camera, even for low-frequency applications, and to invest what often amounts to a great deal more in a pen-and-paper system of limited usefulness. The discussion in the present book reflects this sort of thinking. There is a thorough discussion of paper recorders but the treatment of electronic oscilloscopes does little more than sketch the principles of action and present a photograph of an oscilloscope with directions on how to manipulate the controls. No mention is made of such obvious tricks of the trade as the use of *Z*-modulation to present the signal marker on the same trace as the response, nor the use of stationary spot with modulation in one of the axes only while recording with moving film. The use of an electronic switch in some applications to create a dual-beam oscilloscope from a single

[1] *Laboratory Instrumentation in Psychology* by W. W. Grings. Palo Alto: The National Press, 1954. Pp. 282, $4.75.

beam is not mentioned. Cameras for recording from the oscilloscope rate little more than a line in the text covering recorders. Multibeam oscilloscopes, the relatively new raster-timers, moving film cameras, and similar late developments are not discussed.

This reviewer is also inclined to deplore the space devoted to photographs of electronic gear. Many vastly different circuits may be incorporated into similar housings, and elaborate cabinets in the most gleaming modern tradition frequently contain grossly inferior designs and engineering. The space given to these photographs would be better devoted to tables of equipment sources, performance characteristics, and cost. Particularly obvious in its omission is the mention of the many excellent buys in laboratory equipment in kit form from the Heath Company, a concern which specializes in the manufacture of do-it-yourself electronic equipment of good quality and low cost ideally suited to the usual low budget psychology department.

Similar gaps in useful information exist in discussions other than electronic. For example, in the discussion of visual exposure control no warning is given of the psychologist's *bête noir*—the use of incandescent lamps (other than a few of special design) operated on AC at short exposures. One has only to watch the output of a phototube displayed on the screen of a cathode-ray tube which is monitoring such a source to be impressed with the gross variation in light output correlated with the variation in the 60-cycle current. The contribution of this variation to fluctuation of threshold measures for short durations is considerable.

In view of these defects, and others which will be apparent to readers with varying sorts and amounts of equipment sophistication, it should be emphasized that the book expressly does not intend to be a handbook of instrumentation, but rather an introductory discourse. This reviewer does not wish to be unfair in over-criticizing a book for what it does not do and probably did not intend to do. On the whole the book contains a great number of instrumental ideas, more than a few of which were new to this reviewer, and informative discussions of the characteristics, potential, and use of an enormous variety of apparatus. It is certainly recommended reading along with such works as Strong's *Experimental Physics*, Wilson's *An Introduction to Scientific Research*, Dickinson's *Electrophysiological Technique*, and Whitfield's *An Introduction to Electronics for Physiological Workers*. (*TL*)

There comes a time in the life of every man when he must forget his principles and do what he thinks is right.

—Anonymous

ABSTRACTS OF CURRENT LITERATURE WITH INTERDISCIPLINARY IMPLICATIONS

Prepared by Members of the Staff of the Mental Health Research Institute

Fishman, J. A. **An examination of the process and function of social stereotyping.** *J. soc. Psychol.*, 1956, 43, 27–64.

The author examines the many different conceptions of social stereotyping and attempts to clarify the essential meaning of the construct for social science. The first group of usages treats stereotypes as information contrary to fact. Lippman's original definition, as well as that of Bogardus, Clarke, Centers, and others is included under this rubric. At issue here is the view that stereotypes contain a "kernel of truth." The second group of definitions emphasizes stereotyping as a *process* of inferior judgment. Fishman feels that this is getting at an essential aspect of the construct. He proposes the term *Gleichformierung* for the abridgment of critical thinking that is responsible, in part, for social stereotyping. The third point of view treats stereotypes as the result of attitudinal rigidity. Fishman doubts whether this idea is central, but that the proponents of the view do point out the importance of studying the motivation for social *Gleichformierung*. The last group of definitions emphasizes the saliency of group relatedness as a motivational base for social stereotyping. Fishman sees *perceived* group-relatedness as both the product and a necessary condition for social stereotyping.

The author then considers stereotyping in relation to theories of language and communication. Verbal labeling and other linguistic phenomena are obviously central in the stereotyping behavior, and such concepts as amount of information, redundancy, and recoding, from recent discussions of communication theory may well be useful in a formulation of the stereotype construct.

The next section considers the genesis of stereotypic content and suggests there is a marked similarity between stereotypes and folklore. The theoretical literature on folklore may provide fruitful hypotheses for further research on stereotyping.

The article concludes with a short section presenting a very general paradigm for programs aimed at modifying stereotypic behavior.

(AH)

Küpfmüller, K., and Poklekowski, G. **The regulatory mechanism of volitive motions (Der Regelmechanismus willkürlicher Bewegungen).** *Z. f. Naturforsch.*, 1956, 11b, 1–7.

Subjects were required to transfer a pencil-like object in the shortest possible time from an initial point to a goal point, starting at a given signal (light flash). The two points were marked by metal discs of 1 mm diameter on a table (distance between points: 100 mm, average length of actual trajectory: 121.5 mm). The times t_0 (light flash to lifting of pencil from starting point) and T (light flash to reaching goal) were measured; typical average values in milliseconds (for 11 trials on one subject): $t_0 = 141$, $T = 594$. The light signal was situated near the goal, on which attention was focused immediately before the experiment.

The subjects' performance (intermediate position of pencil vs. time) was recorded photographically and plotted, and thereupon compared with the best possible performance for an ideal linear servomechanism of the same latency time t_0. This theoretical curve showed marked differences from the one obtained in the experiments; in particular the subjects' time T was slightly less than the corresponding time predicted for the ideal servomechanism.

The authors conclude that two regulatory circuits are active in guiding the subjects' movements: 1. An external circuit in which the hand's position is regulated through re-

ports from the eyes to the brain, and 2. an internal circuit, not involving the eyes, but instead proprioceptive impulses from the muscle spindles of the limb. The latter circuit is less accurate but has a smaller latency time and thus allows for greater over-all speed of performance.

According to the authors, most of the motion is internally regulated. The eyes intervene only occasionally to check for major errors; they take over completely at the end, when greater precision is required to hit the goal point. The total motion can, in fact, be subdivided into three main phases, not very sharply distinguished from one another:

(*a*) initial phase (about 20% of total path), used mainly to overcome the inertia of the masses being set in motion;

(*b*) phase of internal regulation (about 75% of total path); this compares favorably with the best possible performance of an ideal linear servomechanism;

(*c*) phase of external regulation, concluding the process.

The internal t_0 (latency time) can be obtained from the experimental curves.

(*ET*)

Lanzetta, J. T. **Group behavior under stress.** *Hum. Relat.*, 1955, 8, 29–52.

How do stressful conditions affect the interaction patterns of a small group? This is the problem investigated in the following experiment.

Twelve four-man groups were each given two types of task, a Reasoning and a Mechanical Assembly task, under three different conditions of increasing stressfulness. The stress was induced by time limits on the solution of the problem, restriction of work space, and badgering of the *Ss* by the *E* about the quality of their performance. In addition, two motivating conditions were set up by having half the groups compete for a prize for the best performance.

The interaction process in the groups was observed and scored according to the system of categories developed by Carter *et al.* In addition, the observers rated, at the end of each session, each group on a number of "group" variables, as morale and cohesiveness, as well as each individual on several individual performance characteristics, as aggressiveness, confidence, etc. The *Ss*, too, gave ratings of their groups at the end of the session.

The principal results were the following: the (rated) quality of task performance was curvilinearly related to amount of stress, the mild stress condition being judged by the observers as most productive. Unfortunately, no objective measures of the task performance were obtained (or reported). In the socio-emotional area of interaction, however, increasing stress produced a *decrease* in negative, tension-producing acts and a corresponding increase in positive, integrating acts. In other words, in the face of stress, imposed largely from the outside, the solidarity of the groups increased although their task performance began to suffer. The two motivating conditions did not produce significant differences; the different types of task, although producing differences in the group activities, did not interact with the stress variable. The limited generality of the results is pointed out.

Although the repeated use of the *same* groups under all conditions on the stress dimension seems desirable from the point of view of statistical analysis, one wonders what effect these replications had on the perception, by the *Ss*, of the stress and its source.

(*FS*)

Lenneberg, E. H. and Roberts, J. M. **The language of experience; a study in methodology.** *Int. J. Amer. Linguistics*, 1956, 22, Memoir 13.

In this monograph, Lenneberg continues his logical and methodological analysis of the Whorf hypothesis as a scientific proposition. (Previous articles are: **"Cognition and ethnolinguistics,"** *Language*, 1953, 29, 463–471 and **"A Note on Cassierer's philosophy of language,"** *Phil. Phenom. Res.*, 1955, 15, 512–522.)

The basic hypothesis in question is that language is expression and/or determinant of the speaker's *Weltanschauung*. Lenneberg sees several possible strategies in approaching the problem experimentally.

The first attack can be made intra-culturally and take the form "linguistics condition *C* is functionally related to non-linguistic behavior *K*." An example is Miller and Selfridge's study of the relation of order of approximation to English word sequences and recall of meaningless strings of words (*Amer. J. Psychol.*, 1950, 63, 176–185). Here all other linguistic and cultural factors are held constant and only the variable in question manipulated.

Another approach is to state that two or more linguistic conditions are related differentially to some nonlinguistic behavior. This can most easily be studied interculturally. Two languages are considered that have some specifiable difference in linguistic structure. Some nonlinguistic phenomenon that is presumably related to the structures in question are then tested for differences in a controlled experimental situation. An example of this approach would be a study of differences in discriminability of isolated sounds that vary with differences in the phonemic boundaries of two languages.

There are a whole group of hypotheses about the relation between language and thought that seem essentially untestable. These are the propositions that refer to unique linguistic phenomenon that occur in only one language. Such a hypothesis would be that there is a relation between language and national character. Such statements tend to refer to subjective data, and to have no comparability between the independent variables from language to language.

Lenneberg proposes three criteria for the selection of data to be used. The first is the universality of the subject matter of the linguistic material. This is what is referred to as the language of experience; those elementary characteristics of the environment that are common to all men. Every one experiences temperature, light, color, time, and many other "basic sensations." The second criterion is that the phenomenon must be variable; that is, manipulable. This is another advantage of the language of experience since methods for handling such material have been thoroughly worked over by psychophysicists. The last criterion is that the material must be simple in order to avoid complication by higher-order cultural factors.

The rest of the monograph is a step-by-step description of the preparation and carrying out of a study of the differences between Zuni and English color categorization. This study, run by J. M. Roberts in Zuni, is a case history of the methodology proposed by Lenneberg.

(*AH*)

Riehl, N. **Electrical conductivity of organic insulators, ice, and proteins in connection with energy transfer. (Elektrische Leitfähigkeit von organischen Isolatoren, Eis und Proteinen im Zusammenhang mit der Energiewanderung).** *Die Naturwissenschaften*, 1956, 43, 145–152.

This paper gives only a brief survey of the author's work, the details of which are published in a Russian journal.

The main problem considered is the transmission of energy inside a (large) protein molecule, so that energy made available at one atomic site of the molecule may be utilized at some other site. The energy transfer itself occurs by displacement of elementary charges (electrons or protons) from site to site. In particular, protons from hydrogen bonds may be so displaced.

Whereas in ordinary dielectric polarization we have a movement of many charges over short distances, the author stresses the importance of a mechanism by which relatively few charges can be transported over large distances. From his studies on the electrical conductivity of organic insulators (e.g. naphthalene), and its temperature dependence, he is led to consider the *formation of "ion pairs."* This occurs when an elementary charge is transferred from one molecule to an immediately adjacent molecule; it is a purely thermal effect and requires less energy than "complete" ionization (charge transported to an "infinite" distance from the original molecule). In the absence of external fields the probability of "recombination", i.e., formation of two neutral molecules from the ion pair, is large. If, however, a strong electric field is applied, the charge may be transported farther away from its original

molecule in a series of "jumps" between adjacent molecules. Thus a kind of "pathway" is established for the elementary charge.

The electrical conductivity, K, of ice for direct current was measured and found to vary with temperature according to the equation $K = K_0\ e^{-B/kT}$ where $B \approx 21$ kcal/mole. A calculation of B based on the assumption that the conductivity is determined by thermal agitation of protons from the hydrogen bonds gave practically the same value. Besides being shifted onto a neighboring "molecule", the protons can change their position around their own oxygen atom, since—in the nonionized state—there are four possible positions to be filled by two protons around each oxyg This "reorientation" of the molecules m be taken into account in calculating B.

Moreover, even in the *absence* of exter fields, the orientations of H_2O dipoles in are such as to *favor* the formation "chains", i.e. pathways of easy transit protons.

The author believes that conditions sim lar to those in ice hold also for the mu more complicated molecules of proteins, pecially if the latter are considered in t naturally occurring state (containing wat molecules in their structure). He gives ge tine, as well as some enzymes, as examp to which his energy transfer mechani could apply.

(E

Books and Pamphlets Received for Review

Barton, R. C. *Our Human Rights*. Washington: Public Affairs Press. 1955. Pp. 102. $2.00.

Beardsley, S. W. and Edgell, A. G. *Human Relations in International Affairs*. Washington: Public Affairs Press. 1956. Pp. 40. $1.00.

Brill, J. *The Chance Character of Human Existence*. New York: Philosophical Library. 1956. Pp. 150. $3.75.

Cohen, J. and Hansel, M. *Risk and Gambling*. New York: Philosophical Library. 1956. Pp. 153. $3.50.

Ewing, U. C. *Thresholds of Existence*. New York: Philosophical Library. 1956. Pp. 286. $3.75.

Fryer, D. H., Feinberg, M. R., Zalkind, S. S. *Developing People in Industry*. New York: Harper. 1956. Pp. 210. $3.50.

Hankin, F. *Making Democracy Work*. Washington: Public Affairs Press. 1956. Pp. 74. $1.00.

Henry, W. E. *The Analysis of Fantasy*. New York: John Wiley. 1956. Pp. 305. $6.00.

Hoch, P. H. and Zubin, J. *Psychopathology of Childhood*. New York: Grune and Stratton. 1955. Pp. 303. $6.00.

Housden, L. *The Prevention of Cruelty to Children*. New York: Philosophical Library. 1956. Pp. 406. $7.50.

Johnson, R. C. *Psychical Research*. New Yor Philosophical Library. 1956. Pp. 176. $2.75.

Lee, R. S. *Psychology and Worship*. New Yor Philosophical Library. 1956. Pp. 110. $3.75.

Mayer, F. and Brower, F. E. *Education for M turity*. Washington: Public Affairs Press. 19 Pp. 155. $3.25.

Rolin, J. *Police Drugs*. New York: Philosophic Library. 1956. Pp. 194. $4.75.

daSilva, Mello, A. *Man, His Life, His Educatio His Happiness*. New York: Philosophi Library. 1956. Pp. 729. $6.00.

Small, L. *Rorschach Location and Scoring Manu* New York: Grune & Stratton. 1956. Pp. 21 $6.50.

Soddy, Kenneth (ed.) *Mental Health and Infa Development*. Vols. I and II. New York: Ba Books. 1956. $4.50 ea.

Stewart, D. A. *Preface to Empathy*. New Yor Philosophical Library. 1956. Pp. 157. $3.75.

Symonds, P. M. *Dynamics of Psychotherapy*. Vo I. New York: Grune and Stratton. 1956. P 211. $5.50.

Zaleznik, A. *Worker Satisfaction and Developme* Boston: Harvard University Division of R search. 1956. Pp. 148. $2.00.